

Cole - #102800 - 6 - 0335080

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by

NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Alan R. Anderson

Kurt Baier

Stephen F. Barker

Monroe Beardsley

Nuel D. Belnap, Jr.

Roderick M. Chisholm

L. Jonathan Cohen

James Collins

James M. Edie

Peter Thomas Geach

Adolf Grünbaum

Carl G. Hempel

John Hospers

Raymond Klibansky

Hugues Leblanc

Ernan McMullin

Benson Mates

José Ferrater Mora

John A. Passmore

Günther Patzig

Richard H. Popkin

Richard Rorty

George A. Schrader

Michael Scriven

Wilfrid Sellars

Alexander Sesonske

Manley H. Thompson, Jr.

James F. Thomson

John W. Yolton

6

VOLUME 6/NUMBER 2

APRIL 1969

CONTENTS

- | | | | |
|--|-----|--|-----|
| I. J. P. DAY: <i>Hope</i> | 89 | VI. WILLIAM A. CHRISTIAN: <i>Religious
Valuations of Scientific Truths.</i> | 144 |
| II. HAIG KHATCHADOURIAN: <i>Objects and
Qualities</i> | 103 | VII. ROLF EBERLE: <i>Universals as Designata
of Predicates</i> | 151 |
| III. ALASTAIR MCKINNON: <i>Kierkegaard's
Pseudonyms: A New Hierarchy</i> | 116 | VIII. DAVID HARRAH: <i>On Completeness in
the Logic of Questions</i> | 158 |
| IV. RICHARD R. K. SORABJI: <i>Aristotle and
Oxford Philosophy</i> | 127 | IX. MARCIA M. EATON: <i>Art, Artifacts,
and Intentions</i> | 165 |
| V. RICHARD TAYLOR: <i>How to Bury the
Mind-Body Problem</i> | 136 | X. STEVEN M. CAHN: <i>The Irrelevance to
Religion of Philosophic Proofs for the
Existence of God</i> | 170 |

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, self-sufficient articles will be published, and not news items, book reviews, critical notices, or "d Qu notes."

MA. Vol. 3

Contributions may be as short as 2,000 words or 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$12, increasing by \$1 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$8, increasing by \$1 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4 per group of 50.

SUBSCRIPTIONS

The price *per annum* is six dollars for individual subscribers and ten dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available, and are sold at the rate of two dollars to individuals, and three dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. These monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.

335080



AMERICAN PHILOSOPHICAL QUARTERLY

Edited by
NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants

Alan R. Anderson
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
Roderick M. Chisholm
L. Jonathan Cohen
James Collins
James M. Edie
José Ferrater-Mora

Peter Thomas Geach
Adolf Grünbaum
Carl G. Hempel
John Hospers
Raymond Klibansky
Hugues Leblanc
Ernan McMullin
Benson Mates
John A. Passmore

Günther Patzig
Richard H. Popkin
Richard Rorty
George A. Schrader
Michael Scriven
Wilfrid Sellars
Alexander Sesonske
Manley H. Thompson, Jr.
James F. Thomson
John W. Yolton



VOLUME 6 (1969)

PUBLISHED BY BASIL BLACKWELL WITH THE COOPERATION OF THE UNIVERSITY OF PITTSBURGH

AMERICAN PHILOSOPHICAL QUARTERLY

CONTENTS OF VOLUME 6 (1969)

	Page
ATWELL, JOHN E. <i>Husserl on Signification and Object</i>	312
BARNHART, J. E. <i>Human Rights as Absolute Claims and Reasonable Expectations</i>	335
BEARDSLEY, ELIZABETH L. <i>A Plea for Deserts</i>	33
BROYLES, JAMES E. <i>Language and Common Sense</i>	233
CAHN, STEVEN M. <i>The Irrelevance to Religion of Philosophic Proofs for the Existence of God</i>	170
CASTANEDA, HECTOR-NERI <i>Ought, Value, and Utilitarianism</i>	257
CHANDLER, HUGH S. <i>Butler on Bodies</i>	84
CHRISTIAN, WILLIAM A. <i>Religious Valuations of Scientific Truths</i>	144
<i>Corrigenda</i>	341
COWAN, J. L. <i>Deliberation and Determinism</i>	53
CUA, ANTONIO S. <i>Morality and the Paradigmatic Individuals</i>	324
CUNNINGHAM, G. WATTS <i>On Reason's Reach: Historical Observations</i>	1
DANIELS, CHARLES B. <i>Personal Identity</i>	226
DAY, J. P. <i>Hope</i>	89
DICKIE, GEORGE <i>Defining Art</i>	253
EATON, MARCIA M. <i>Art, Artifacts, and Intentions</i>	165
EBERLE, ROLF <i>Universals as Designata of Predicates</i>	151
GROSSMANN, REINHARDT <i>Non-Existent Objects: Recent Work on Brentano and Meinong</i>	17
HARRAH, DAVID <i>On Completeness in the Logic of Questions</i>	158
KHATCHADOURIAN, HAIG <i>Objects and Qualities</i>	103
KÜNG, GUIDO <i>The Role of Language in Phenomenological Analysis</i>	330
LANDESMAN, CHARLES <i>Actions as Universals: An Inquiry into the Metaphysics of Action</i>	247
LYCAN, W. GREGORY <i>On Intentionality and the Psychological</i>	305
LYONS, DAVID <i>Rights, Claimants, and Beneficiaries</i>	173
MAUTNER, THOMAS <i>Aquinas's Third Way</i>	298
MCKINNON, ALASTAIR <i>Kierkegaard's Pseudonyms: A New Hierarchy</i>	116
MUNSAT, STANLEY <i>What Is A Process?</i>	79
OWENS, MEIRLYS <i>The Notion of Human Rights: A Reconsideration</i>	240
PIKE, NELSON <i>Omnipotence and God's Ability to Sin</i>	208
PITCHER, GEORGE <i>Minds and Ideas in Berkeley</i>	198
PRIOR, ARTHUR N. <i>Tensed Propositions as Predicates</i>	290
RICHMAN, ROBERT J. <i>Responsibility and the Causation of Actions</i>	186
SARTORIUS, ROLF <i>A Neglected Aspect of the Relationship Between Berkeley's Theory of Vision and His Immaterialism</i>	318
SHERIDAN, GREGORY <i>The Electroencephalogram Argument Against In corrigibility</i>	62
SIMON, MICHAEL A. <i>Could There Be A Conscious Automaton?</i>	71
SLOMAN, AARON <i>How to Derive "Better" From "Is"</i>	43
SORABJI, RICHARD R. K. <i>Aristotle and Oxford Philosophy</i>	127
STOCKER, MICHAEL <i>Consequentialism and Its Complexities</i>	276
TAYLOR, RICHARD <i>How to Bury the Mind-Body Problem</i>	136
WALKER, JEREMY D. B. <i>Statements and Performatives</i>	217

I. HOPE

J. P. DAY

I. INTRODUCTION

PHILOSOPHERS seem to have abandoned Hope. This is surprising, since Hope is an interesting concept, as I shall hope to show; as is evidenced, moreover, by the fact that in the past four great philosophers have taken it very seriously. Aquinas regards Hope as the second Theological Virtue. Hume finds Hope and its opposite, Fear, to be the most important of what he calls "the direct passions." Kant tells us that "What may I hope if I do my duty?" is one of the fundamental questions of philosophy. Similarly to Kant, Mill maintains that the fundamental question in religion is not "What is it legitimate to believe?", but "What is it legitimate to hope?"

This paper falls into two parts, a critical and a constructive. In the former part we shall consider three questions: (1) Is Hope an emotion? This will be devoted to a critical exposition of the views of Hume and of his successors, the experimental psychologists. (2) Is Hope a right? In this connection, we shall consider the opinions of Kant and Mill. (3) Is Hope a duty? Here, the doctrine of Aquinas and of his successors, the Christian moral theologians, will be discussed. In the latter, constructive, part of the paper I shall give my own answer to these three questions, making due use of the conclusions arrived at in the former, critical part. Two additional questions will also be answered, namely, (4) what does "hope" mean?; and (5) what are the resemblances and differences between Hope and Belief?

II. CRITICAL EXPOSITION

Is Hope an Emotion?

Hume and the Experimental Psychologists

The gist of Hume's¹ doctrine is as follows: Hope and Fear are "direct passions" (i.e. direct emotions). The other direct passions are Desire and Aversion, Joy and Grief. Hope and Fear are opposite or contrary emotions; as are also Desire and Aversion, Joy and Grief. Furthermore, he adds, Hope and

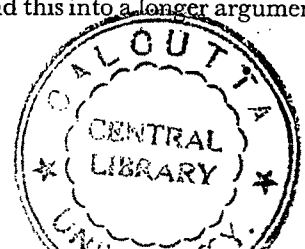
Fear are mixtures of Joy and Grief. For opposite emotions can mix, or alternate, or destroy each other. When an event, *E*, is good and certain, a person, *A*, feels joy at the prospect of *E*. When, however, *E* is bad and certain, *A* feels grief at the prospect of *E*. When *E* is good and probable, *A* hopes that *E* will occur. Finally, when *E* is bad and probable, *A* fears that *E* will occur. Whether *A* hopes that *E* will occur or fears that *E* will occur depends on whether joy or grief predominates in the mixture. This depends in turn on two factors, namely, (1) whether *E* is good or bad, and (2) whether *E* is probable or improbable. For example, as just said, when *E* is good and probable, *A* hopes for *E*.

I disagree with Hume that Hope is an emotion, notwithstanding that his opinion is shared by virtually all philosophers (including Aristotle, Aquinas, and Descartes) and by many, perhaps most, psychologists. Here, first, is a short argument in support of this position. Hope involves (1) desiring and (2) estimating a probability. In other words, "*A* hopes that *P*" is true if and only if "*A* wishes that *P*, and *A* thinks that *P* has some degree of probability, however small" is true. (*P* is, of course, a proposition-variable.) Let us call these two constituents of Hope the desiderative constituent and the estimative constituent respectively. Now, the estimation of probabilities is plainly not an emotion. But, as Antony Kenny² has pointed out, neither is Desire, despite the fact that it is often alleged to be so—by Hume himself, for instance, as we have just seen. The reason is that Desire normally lacks two of three constituents thought to be necessary to emotion; namely, some characteristic sensation and some characteristic physical symptom. To be sure, it is sometimes the case that when *A* wishes that *P* he feels a yearning in his bowels and he also pants. But generally this is not so, and when *A* says "I wish that *P*" he is not doing either of these things. But since a hope is identical with a desire plus a probability-estimate, and neither desires nor probability-estimates are emotions, Hope is not an emotion either.

But now to expand this into a longer argument. A

¹ D. Hume, *A Treatise of Human Nature*, Bk. II, Pt. III, sect. IX.

² A. Kenny, *Action, Emotion and Will* (London and New York, 1963), chs. II, V, IX.



genuine emotion, such as Fear, is a complex type of occurrence. It is generally agreed that, for "*A* fears *X*" to be true, the following conditions, at least, must be satisfied. (1) *A* has a sensation characteristic of that emotion—say, a sinking feeling in the pit of his stomach. (2) He exhibits some physical symptom characteristic of that emotion—say, he pales. (3) He displays some behavior-pattern characteristic of that emotion—for instance, he flees, or takes some other avoiding action. Now, the claim of Hope to be an emotion rests on the fact that it involves Desire. But, as we have just seen, Desire's own claim to be an emotion is unacceptable because normally the first two conditions are not satisfied when *A* wishes that *P*. But what about condition (3)? There is indeed a behavior-pattern characteristic of Desire, and hence of Hope. For if *A* wishes that *P*, then he will be disposed to try to bring it about that *P*. Evidently, however, this will hold true only in those cases where *A* can try to bring it about that *P*, which are those cases where the object of hope is an action or passion of the subject. These cases, however, form a minority of the total of any man's hopes. Thus, compare the four following statements: (1) "*A* hopes that he (*A*) will win the race"; (2) "*A* hopes that he (*A*) will be elected"; (3) "*A* hopes that *B* will win the race"; and (4) "*A* hopes that the sun will shine tomorrow." *A* can try to bring it about that *P* in cases (1) and (2), but not in cases (3) and (4). Linguistically, this is reflected in the fact that cases (1) and (2) can be expressed by a "hope-to" sentence-construction, whereas (3) and (4) cannot. E.g., "*A* hopes to win the race." So generally, there can be a proposition of the form "*A* hopes to *V*" corresponding to a proposition of the form "*A* hopes that *P*" only when '*P*' is equivalent to "*A* will *V*"; i.e., when the subject of the second verb is identical with the subject of the first verb, namely, "hopes." In other words, "*A* hopes to *V*" is equivalent to "*A* hopes that *A* will *V*," but not to "*A* hopes that *B* (or *X*) will *V*." (*V* is a deed-variable, and *X* is a thing-variable.) The upshot of these considerations is, therefore, that the three chief tests of genuine emotion—viz., characteristic sensation, characteristic physical symptom, and characteristic behavior-pattern—are satisfied in the case of Hope much less often than not. It therefore seems proper to withhold the name "emotion" from Hope.

Next, it will be informative to speculate on the probable reasons for the contrary opinion that Hope is an emotion.

First, of course, comes the fact that some or all of the three main tests are sometimes satisfied, even if in a minority of cases.

Secondly, Hope has two important properties which genuine emotions also have. It has an intentional object, as Brentano terms it. That is, as *A* is afraid of *X* and angry with *B*, so he hopes for *E*. Again, Hope can be a motive for action. Corresponding to "*A* *V*s because *A* fears *X*" we have "*A* *V*s because *A* hopes that *P*." E.g., respectively: "Jane deserted her husband because she was frightened of him" and "Jim backed the grey mare because he hoped to make a lot of money thereby." Thirdly, because Hope includes Desire, and Desire is commonly, though mistakenly, thought to be an emotion. It must be pointed out, however, that even if Desire were an emotion, the consideration just advanced would show at best that Hope involves emotion, not that it is an emotion. For the other constituent of Hope, the estimative one, is decidedly not an emotion. Fourthly, the fact that the opposite (or more accurately, one contrary) of Hope is Fear. This introduces a point of interest about Fear.

The argument goes, presumably, like this: The opposite of any emotion must itself be an emotion. The opposite of Hope is Fear. Fear is an emotion. Therefore, Hope is an emotion. However, the argument is invalid because the term "fear" has different meanings in the second and third premisses. For the sense of "fear" in which Fear is a contrary of Hope is not the sense of "fear" in which Fear is an emotion. For one must distinguish "*A* fears death" (and, generally, "*A* fears *X*") from "*A* fears that he is dying" (and, generally, "*A* fears that *P*"). It is "*A* fears that *P*" which is contrary to "*A* hopes that *P*," but it is "*A* fears *X*" which ascribes an emotion to *A*. Nevertheless, these two uses of the word "fear" are connected, which is perhaps the reason why the same word is used for both concepts. For if a man fears death then he will, in certain circumstances, fear that he is dying—as opposed to, say, hope that he is dying, or be indifferent whether (i.e., neither hope nor fear that) he is dying. This diagnosis applies to Hume, for he certainly fails to distinguish between fearing *X* and fearing that *P*. More will be said later about the opposites of Hope (in Sect. 2.3), when we shall see that Hope has, in fact, at least two important contraries, Fear and Despair. The question of the opposition of emotions was first treated by Aristotle.³ "Opposite" must be taken in the same sense as it has in the Square of Opposition.

³ Aristotle, *Rhetorica*, tr. by W. R. Roberts, and ed. by W. D. Ross (Oxford, 1924), Bk. II, esp. 1377 b 13—1378 a 31, 1382 a 19—1383 b 11.

Thus, Hope and Fear are not contradictories because they are not exhaustive alternatives; as we have just seen, *A* neither hopes nor fears that *P* when he could not care less whether or not *P*. But they are contraries, or exclusive alternatives, since *A* cannot simultaneously hope and fear that *P*. He can, however, successively hope and fear that *P*; but this is, in Hume's terminology, "alternation" not "mixture" of passions. For instance, if *A* has applied for a certain post, but is uncertain whether he really wants it or not, he is likely to oscillate between hoping and fearing that he will be offered it. According to Aristotle, the opposite of Fear the emotion is Confidence, or, in more modern terminology, the sense of security. It is a nice question whether Fear and Confidence are contraries or contradictories; I myself incline to the latter opinion.

Fifthly, it seems likely that another reason why Hume thinks that Hope is an emotion is his opinion that Hope is a mixture of Joy and Grief. For the argument here goes, presumably, thus: a mixture of contrary emotions is itself an emotion; Joy and Grief are contrary emotions; Hope is a mixture of Joy and Grief; therefore, Hope is an emotion. Now, Hume is certainly right in pointing out that opposed emotions can mix. For example, a mother may have "mixed feelings" at her daughter's wedding; she is glad at her daughter's happiness, but sad at losing her. Moreover, where there are mixed feelings, there are also likely to be mixed physical symptoms; the mother in the example may well be smiling through her tears. But mixed Joy and Grief are surely not Hope or Fear, as Hume alleges. The mixed feeling of the mother in the example is certainly neither Hope nor Fear.

Lastly, Hope resembles genuine emotions in admitting of degrees. We shall revert to this important feature of Hope shortly.

Hume's main object is to provide a psychological analysis of Hope and Fear. The core of this is: "It is a probable good or evil that commonly produces hope or fear." Or, Hope and Fear are produced by mixing Joy and Grief much as, to use Hume's own simile, salad dressing is produced by mixing oil and vinegar. But as is often the case with Hume, he provides a philosophical analysis as well as a psychological one. In the case of Hope, this is his assertion that "*A* hopes for *E*" is true if and only if "*E* is desirable (or good) and probable" is true. But this is a mistake. The analysis should read rather: "*A* hopes for *E*"

is true if and only if "*E* is desired by *A* and thought probable by *A*." Nevertheless, Hume deserves credit for seeing that the meaning of "Hope" must be explained in terms of "Desire" and "Probability."

Kenny has pointed out that modern experimental psychologists, when investigating the emotions, have taken over a number of the basic doctrines of those two influential treatises on the passions, Descartes' and Hume's. It comes as no surprise, therefore, to find them too asserting that Hope is an emotion. J. Drever,⁴ e.g., tells us that Hope is

best regarded as an emotional attitude, though designated by some psychologists a derived emotion. Its predominant characteristic is desire for the attainment of some objective, with some idea that this desire will be satisfied, giving a pleasant hedonic tone to the experience.

Some call Hope a "prospective" emotion on the ground that the object of Hope is always future. But this seems wrong, since one can perfectly well say, e.g., "I hope that it is not freezing outside" and "I hope that you slept well"; where the objects of hope are respectively present and past.

The emotions, we have seen, admit of degrees. One way in which experimental psychologists have investigated the emotions has been by attempting to measure their intensities. This is usually done by measuring one constituent of the emotion-syndrome, viz., the physical symptoms. Thus, Fear has been measured by recording, among other things, changes in the rate and force of the subject's heart-beat on an electrocardiograph. Now, since the psychologists⁵ regard Hope as an emotion, and since there are in fact degrees of Hope, they have attempted to measure Hope also. However, they have found that, in Kenny's words, Hope is "not patient of this type of investigation." Nor, in the light of what has already been said, is it hard to see why. For we have noticed that normally Desire, and hence Hope, do not involve any physical symptom, such as panting, at all. So that any attempt to measure Hope by, say, recording the rate of the subject's breathing on a pneumograph is bound to meet with very limited success.

But let us now consider this question from the linguistic standpoint. That there are indeed degrees of Hope is revealed clearly by our ways of speech. But if we attend carefully to the relevant locutions, we come upon an interesting fact. Namely, that

⁴ J. Drever, *Dictionary of Psychology* (London, 1964, revised ed.).

⁵ See, e.g., J. M. Baldwin and G. F. Stout, "Hope," *Dictionary of Philosophy and Psychology*, ed. by J. M. Baldwin (London and New York, 1901).

degrees of Hope are expressed in two different ways. On the one hand we say, for example, "A has high hopes that P," or "A has only a faint hope that P." On the other hand we say, for example, "A hopes fervently that P," or "A hopes, but without much enthusiasm, that P." Plainly, the "high"/"faint" scale relates to the estimative constituent of Hope, whereas the "fervent"/"tepid" scale relates to the desiderative constituent of Hope. For there are, of course, degrees both of Probability and of Desire. Moreover, if these two constituents of Hope could be measured, in some sense of "measure," at all, they would have to be measured in different ways. The degree of intensity if A's desire that P may be determined, at least sometimes, by measuring the rate of his breathing, or by asking him how badly he wants that P, or by observing how hard he tries to bring it about that P, or by noting how much money he is prepared to pay in order to secure that P. But the degree of probability which A attaches to P will have to be determined by, say, asking him how probable he thinks P is, or, perhaps, by noting the odds at which he is prepared to bet on P.

Is Hope a Right? Kant and Mill

Kant⁶ maintains that the question "What may I hope?" ranks with the questions "What can I know?" and "What ought I to do?" as one of the fundamental problems of philosophy. He adds that the peculiarity of the question is that it is both theoretical and practical.

In answer to the question, Kant says that one may indulge two important hopes. First, the hope that happiness is proportioned to virtue. This is because it is possible that there is a noumenal world which is governed by a moral Being Who has sufficient power so to proportion happiness to virtue. Secondly, the hope that the soul is immortal. In his own words: "On the basis of his previous progress from the worse to the morally better, and of the immutability of intention which thus becomes known to him, one may hope for a further uninterrupted continuance of this progress, however long his existence may last, even beyond this life." As L. W. Beck⁷ has pointed out, although Kant calls the immortality of the soul a "postulate of pure practical reason," the so-

called "postulate" turns out to be in fact neither more nor less than a hope.

Mill⁸ argues along lines which seem to be unmistakably Kantian. His main thesis is that it is "legitimate" for A to hope that P when there is no evidence relevant to P (i.e., neither for P nor against P). Naturally, Mill is concerned here with what he calls "supernatural hopes," and in particular with the hopes that God exists and that there is a life after death. Mill defends these supernatural hopes on the score of their beneficial effects on the hoper. He also observes that "a hopeful disposition gives a spur to the faculties and keeps all the active energies in good working order."

Richard Taylor⁹ points out that Mill's main thesis about Hope was subsequently taken up by James in his essay "The Will to Believe." James,¹⁰ incidentally, wanted to change the title of this essay to "The Right to Believe," which is certainly more apt. For he is concerned in it with the same sort of questions about Belief as we are concerned with in this section of this paper, which is entitled "Is Hope a Right?"

Finally, Mill represents hoping as a sort of imagining, and lays down "the principle that in the regulation of the imagination, literal truth of facts is not the only thing to be considered. Truth is the province of reason . . ."

Now to comment on these doctrines, beginning with Mill's contention that hoping is a sort of imagining. It seems to be clearly false, since neither desiring nor estimating probabilities are imagining. Nevertheless, there is a connection of some interest between Hope and Imagination, namely this. In the case of some hopes, of which Mill's "supernatural hopes" are a conspicuous example, it is necessary to be able to imagine the object of hope to be able to want it, and so to hope for it. Thus, one cannot hope for immortality, or for a society in which happiness is proportioned to virtue, unless he is able to imagine what these states of affairs would be like. However, most hopes are not like this. I need no imagination to hope that the sun will shine tomorrow because I know from previous experience what sunshine is like, and so have no need to imagine what it would be like.

We have seen that Mill defends both supernatural hopes and a hopeful disposition on the score of their

⁶ I. Kant, *Critique of Pure Reason*, tr. by N. K. Smith (London, 1933), B 833—B 837; *Critique of Practical Reason*, tr. and ed. with Introduction by L. W. Beck (Chicago, 1949), Bk. II, ch. II, sects. II—V.

⁷ L. W. Beck, *A Commentary on Kant's Critique of Practical Reason* (Chicago, 1960), p. 270.

⁸ J. S. Mill, *Theism* (1875, 4th ed.), ed. with Introduction by R. Taylor (New York, 1957), esp. Pt. V.

⁹ See note 8.

¹⁰ William James, *Selected Papers on Philosophy*, ed. with Introduction by C. M. Bakewell (London, 1917), p. 99. Note.

beneficial effects on the hoper. It would be irrelevant to dispute with Mill about whether the hope for immortality, for example, or hopefulness are or are not beneficial to the hoper. It is relevant, however, to point out that we have here a significant ambiguity in the word "hope." When it is said, for instance, that Hope is a good thing, this may mean either that hopefulness, i.e., Optimism or a hopeful disposition, is a good thing; or alternatively, that some particular hope, for instance, the hope for immortality is a good thing. Aquinas and the moral theologians, Kant and Mill, are all concerned with particular hopes, especially supernatural ones, not with hopefulness; and similarly there is nothing in this paper about Optimism.¹¹

Now for Mill's main thesis, that if there is no evidence relevant to *P*, then *A* has a right to hope that *P*. The chief criticism of it is as follows. *A* has a right to hope that *P* if and only if his hope that *P* is (1) neither immoral nor imprudent, and (2) reasonable. Further, *A*'s hope that *P* is reasonable if and only if (1) *P* is probable in some degree, however small; (2) the degree of *A*'s probability-estimate that *P* corresponds to the degree of probability that *P*; and (3) *P* is consistent both with itself and with the other objects of *A*'s hopes, *Q*, *R*, etc. But condition (1) presupposes, in P. F. Strawson's¹² sense of "presuppose," that there is evidence relevant to *P*. Hence, *A*'s hope that *P* is neither reasonable nor, consequently, legitimate when there is no evidence relevant to *P*.

There can be no evidence relevant to *P* in two different ways. First, there is the case where such evidence simply does not exist. Secondly, there is the case where such evidence exists in a sense, but is self-destructive. To illustrate, suppose that the evidence consists of two pieces, *Q* and *R*, and that *Q* makes it very probable that *P*, whereas *R* makes it very probable that not-*P*. It is worth while to draw attention to the second sort of case, since it is sometimes argued that one has a right to hope for, say, immortality, because the arguments or evidences *pro* have the same probative force as the arguments or evidences *con*. But the objection to this line of argument is, of course, exactly the same as that which has just been directed against Mill's main thesis.

But now, what sort of a right is the right to hope, and in particular is it ever a moral right? The answer seems to be that the right is never purely moral, but that it may be a mixed moral and prudential right. The latter case arises when *A*'s hope that *P* is (1) not immoral and (2) reasonable. The moral constituent of the right derives from (1) and the prudential constituent from (2). Hopes, therefore, can be appraised morally and/or prudentially. The prudential appraisal attaches to the estimative element in Hope, and the moral appraisal attaches to the desiderative element in Hope. One can perfectly well say, for example, "*A*'s hope that *B* will be killed is reasonable but immoral," the immorality consisting in the fact that *A* is manifesting a bad will, as Kant would say. That hopes can be appraised morally and/or prudentially is one of the important differences between them and beliefs; for I agree with H. H. Price¹³ that beliefs can only be appraised prudentially.

The right to make a given probability-estimate is prudential for the following reason. "*A* has a right to think *P* very probable" means "given the evidence that he has, *A* is being reasonable in thinking *P* very probable." But the right to hope reasonably (which means, principally, the right to proportion the degree of one's probability-estimate of *P* to the degree of probability of *P*) is a prudential right, because it is to our interest to do so. Similarly, the right to believe reasonably (which means, principally, the right to proportion the degree of one's belief that *P* to the degree of probability of *P*) is also a prudential right for the same reason.

Finally, these considerations show that Kant is right in saying that the question "What may I hope?"—or better, the question "When may I hope that *P*?"—is both theoretical and practical. It is theoretical, because the question "Has *A* a right to hope that *P*?" entails the question "Is *A*'s probability-estimate of *P* reasonable?" But it is also practical, because it also entails the question "Is *A*'s wish that *P* neither immoral nor imprudent?"

Is Hope a Duty?

Aquinas and the Moral Theologians

Aquinas¹⁴ discusses Hope in two separate but connected contexts; first, in that part of his treatise

¹¹ For a recent discussion of Optimism, see M. A. Boden, "Optimism," *Philosophy*, vol. 41 (1966), pp. 291-303.

¹² P. F. Strawson, *Introduction to Logical Theory* (London and New York, 1952), ch. 6, Pt. III, sect. 7.

¹³ H. H. Price, "Belief and Action," *Proceedings of the Aristotelian Society*, Supplementary Volume XXVIII (1954), pp. 1-26; "Half-Belief," *ibid.*, vol. XXXVIII (1964), pp. 149-162.

¹⁴ Thomas Aquinas, *Summa Theologiae*, tr. with Intro. by J. P. Reid, Vol. XXI (London and New York, 1965), 1a2ae. 40, 1-1a2ae. 40, 8; *Summa Theologica*, tr. by Fathers of the English Dominican Province, Pt. II, Second Part, 9, First Number (London, 1916), Questions XVII-XXII.

on the passions which concerns Fear and Anger; and secondly, in that part of his treatise on the virtues which concerns the Theological Virtues, which are Faith, Hope, and Love: Aquinas derives this list of Theological Virtues, of course, from St. Paul.¹⁵

Aquinas' opinions on Hope considered as an emotion are as follows. Hope ranks with Fear, Joy, and Grief as one of the principal emotions. He defines "Hope" thus: "Hope is a movement of appetite aroused by the perception of what is agreeable, future, arduous, and possible of attainment."

Aquinas tells us further that Hope involves both cognition and appetite, since *A* must perceive *X* before he can desire it. The sort of cognition involved is sometimes called "expectation," in the Latin sense of "watching out for." Again, Aquinas writes: "The possible which is the object of hope is that which refers to the scope of an agent." That is, first, it is not the logically possible or non-self-contradictory. Secondly, there are two species of this possible: (1) that which can be done by Man, and (2) that which can be done by God. The second species of possibility is connected with Hope considered, not as an emotion, but as a Theological Virtue. Aquinas goes on to point out that Hope has two contraries, Fear and Despair, and that these two contrarieties arise for different reasons. "Fear and Hope are contraries by reason of their objects which are, respectively, disagreeable and agreeable." "But the only contrariety between Hope and Despair is that of approach and withdrawal." The reason why there is withdrawal in the case of Despair is that, although the object is still desired, it is regarded as unattainable. Aquinas observes further that Hope is an aid to action in two ways. First, because the subject's awareness of the difficulty makes him concentrate his efforts on surmounting it; and secondly, because Hope causes Delight, which makes for more effective operation.

Let us now turn to Aquinas' views on Hope considered as a Theological Virtue. It is to be noted, in the first place, that Aquinas is discussing here one particular hope, namely, Man's hope of attaining, with God's help, eternal happiness in the form of union with God. The clause "with God's help" involves that the object of this hope is possible in the second sense distinguished above. Aquinas says explicitly that we ought to hope for eternal happiness, and he identifies enjoying eternal happiness with enjoying God Himself.

It follows from the last point that this hope is a

Theological Virtue, since "a Theological Virtue" means "one that has God for its object." This hope for eternal happiness, Aquinas continues, is not an emotion but a habit of the mind. For Man has a higher and a lower appetite, the sensitive and the intellective, the name of the intellective appetite being the "Will." Now, this hope resides in the Will. Lastly, Aquinas informs us that "This habit . . . does not flow from our merits, but from grace alone."

Now for some critical comments on Aquinas' theses, beginning with Fear and Despair as contraries of Hope. Aquinas is certainly correct in saying this. For *A* cannot despair that *P* at the same time as he hopes that *P*, any more than he can fear that *P* at the same time as he hopes that *P*. He is also correct in drawing the main distinction between Fear and Despair as he does; namely, by pointing out that "*A* fears that *P*" entails "*A* wishes that not-*P*," whereas "*A* despairs that *P*" entails "*A* wishes that *P*." Thus, Fear must differ from Hope in respect of the desiderative element; for *P* is wanted in Hope, but not-*P* is wanted in Fear. But Fear need not differ from Hope in respect to the estimative element; for the same degree of probability may be ascribed to *P* in both cases. By contrast, Despair must differ from Hope in respect to the estimative element; since *P* is ascribed some probability in Hope, but none in Despair. But Despair need not differ from Hope in respect to the desiderative element; for *P* may be wanted in the same degree in both Hope and Despair.

To turn now to Aquinas' thesis that the Christian hope for eternal happiness is a habit which resides in the Will. By "habit" Aquinas presumably means "disposition." This, too, is true, since "*A* hopes that that *P*" does not describe an actually occurrent process in the way that "*A* is washing his hands" does. Rather it means that, if *Q*, then *A* would *V*. E.g., if he were asked whether he wishes that *P*, he would answer affirmatively. It is of interest that Bain¹⁶ agrees with Aquinas both that Hope is a disposition, and that Hope resides in the Will. For on Bain's view, Hope is a species of Belief, and Belief is a disposition or tendency, the only test of the presence of which is action. Hence, Belief (and therefore Hope) relate to action and the Will, not to passion, as Hume and the experimental psychologists maintain. However, Bain seems to overlook the estimative element in hoping. Certainly, hoping necessarily involves wishing or willing; but it is not identical with it. Also, Bain is quite wrong in repre-

¹⁵ I *Corinthians*, xiii, 13.

¹⁶ Alexander Bain, *The Emotions and the Will* (London, 1859), see the chapters on "The Will" and on "Belief."

senting Hope as a species of Belief. The differences between them will be indicated later.

Aquinas' thesis that Hope is identical with Expectation plus Desire is apparently identical with the definition of "Hope" in the Oxford Dictionary. But it is not so really. For by "expect" Aquinas means "watch out for," and his view is that *A* does this because *A* must perceive *X* before he can desire *X*, and must desire *X* before he can hope for it. (This opinion, incidentally, is surely refuted by Mill's point that the object of a man's hope is sometimes something which is wholly imaginary, the like of which he has never perceived before.) By contrast, the dictionary means by "expectation" something like what is called in this paper "estimation of probabilities." Nevertheless, the view that "*A* hopes that *P*" means "*A* wishes that *P* and thinks that *P* is in some degree probable" is not identical with the dictionary definition, according to which it means "*A* wishes and expects that *P*." This is because "*A* expects that *P*" is not synonymous with "*A* thinks that *P* is probable in some degree, however small," but rather with "*A* thinks that *P* is more probable than not." In other words, "*A* expects that *P*" is synonymous with "*A* believes that *P*." Incidentally, there is a different and secondary use of "expect." "England expects that every man this day will do his duty" does not mean "Englishmen believe that every sailor in the Fleet will do his duty today," but means rather "Englishmen are relying, or counting, on every sailor in the Fleet doing his duty today."

Aquinas' thesis that the object of hope must be possible raises the important question: is the object of hope that which the subject of hope thinks to be probable (as I maintain) or that which he thinks to be possible? Notice that Aquinas says that the object of hope must be possible, not be thought by the subject of hope to be possible. The difference is, of course, vital. But it is the thesis that the object of hope is that which the subject thinks to be possible which is the interesting one.

It will be convenient to begin by developing my own thesis further. By "probable" I mean "probable in some degree, however small (e.g. 1/100)." This use is therefore wider than the usual one, since in common speech "probable" means "more probable than not (or greater than 1/2)." In other words, whereas in normal use "*P* is probable" and "*P* is improbable" are contraries, in this use they are not.

Now, "*P* is probable" (in my use) is also contrary to all of the following: "*P* is analytic," "*P* is self-contradictory," "*P* is true," "*P* is false," "*P* is

certain," and "*P* is contracertain." (By "*P* is contracertain" is meant "not-*P* is certain.") It will be seen from the last two cases that my use of "probable" excludes "probable in the degree 1 (i.e. certain)" and "probable in the degree 0 (i.e., contra-certain)." In saying, therefore, that the object of hope is that which the subject thinks to be probable, I am also saying by implication that the object of hope is not that which the subject thinks to be analytic, self-contradictory, true, false, certain, contracertain. To illustrate, consider two examples, the self-contradictory and the certain. It makes no sense to say, e.g., "*A* hopes to square the circle tomorrow although he thinks it cannot be done," or "*A* hopes to die before his 1,000th birthday although he thinks he certainly will." In making a philosophical analysis of "Hope," it is essential to realize that there are restrictions on the possible objects of hope; or, in other words, that it is by no means the case that, in the formula "*A* hopes that *P*," any proposition whatever can be substituted on the variable *P*. The seemingly widespread opinion that one can hope for anything is, in fact, very wide of the mark.

Naturally, a similar question can be asked about the subject of hope. The short answer seems to be that, in the formula "*A* hopes that *P*," *A* can only be some person; though the question can certainly be raised whether it makes sense to say, for example, that a dog is hoping to be given a bone. Aquinas does, in fact, ask whether animals can hope, and answers that they can. It is interesting to observe, in this connection, that Aquinas contends that it is by no means the case that all men can hope to attain eternal happiness in the form of union with God. It is impossible for two large classes of men, namely, the blessed and the damned. His reasons for these exclusions are logically impeccable. For the blessed are, by definition, those who have already attained eternal happiness, and know it. But one cannot, logically, want, and so hope for, what he already has, and knows that he has. Similarly, the damned are, by definition, those who cannot attain eternal happiness, and know it. But one cannot, logically, want, and so hope for, what he cannot have, and knows that he cannot have. That is why, in Dante's¹⁷ *Divine Comedy*, there is written above the Gate of Hell:

All hope abandon, ye who enter here.

Of course, both the blessed and the damned could hope for eternal happiness if they did not know that they have it, and cannot have it, respectively; or if they thought, mistakenly, that they did not have it,

¹⁷ Dante Alighieri, *Inferno*, Canto III, line 9.

and can have it, respectively. If this were so, the blessed would be in the position of those malcontents who, as the saying goes, do not know when they are well off.

Let us consider now the rival thesis that the object of hope is that which the subject thinks to be possible. For a start, it must be pointed out that there is a sense of "possible" in which this thesis is only apparently, and is not really, opposed to my own thesis. For, as William Kneale¹⁸ has remarked, there is a use of "Possibility" in which it is synonymous with "Probability." E.g., "there is a distinct possibility that *P*" means the same thing as "there is a fair probability that *P*." Hence, too, the "classical" theorists of Probability, J. Bernoulli and Laplace, speak indifferently of "equally probable alternatives" and of "equally possible alternatives."

Usually, however, "possible" means "logically and physically possible"; the meaning of which, in turn, is "not self-contradictory, not false, and not contraccertain." Then, the objection to the thesis that the object of hope is that which the subject thinks to be possible, in this sense of "possible," is that it is too wide. For "*P* is possible," unlike "*P* is probable," is compatible with "*P* is analytic," with "*P* is true," and with "*P* is certain." The thesis, therefore, admits as objects of hope things which cannot be so. For all the following statements are clearly wrong: "*A* hopes that all bachelors are unmarried although he thinks that the proposition 'all bachelors are unmarried' is analytic"; "*A* hopes that today is Saturday although he thinks that the proposition 'today is Saturday' is true"; and (as already said) "*A* hopes that he will die before his 1,000th birthday although he thinks that the proposition '*A* will die before his 1,000th birthday' is certain."

Finally, some comments on Aquinas' theses that "we ought to hope for . . . eternal happiness" and that "(this happiness) . . . does not flow from our merits, but from grace alone." In other words, let us consider the question: is there ever a duty to hope, and if so, when? Some modern moral theologians agree with Aquinas that the answer to this question is affirmative. We shall take up first an objection which has been brought against them by T. H. McPherson.¹⁹ They say that there are seven Christian virtues, comprising three "Theological" ones (namely, Faith, Hope, and Love) and four "Cardinal" ones (namely, Prudence, Fortitude,

Temperance, and Justice). They state further that the difference between the Theological Virtues and the Cardinal ones is that the former are "infused" (i.e., implanted in us by the grace of God) whereas the latter are acquired by practice. But if *A* can only hope for eternal happiness if God has implanted that hope in him, it cannot be right to say, as Aquinas does, that *A* ought to hope for eternal happiness; for "ought" implies "can."

If "Theological Virtue" is defined as "infused virtue," McPherson's objection is surely decisive. However, it is worth remarking that, in his discussion of the Christian hope for eternal happiness, Aquinas defines "Theological Virtue" differently, as we have seen; namely, as "a virtue which has God for its object." Thus, Love is love of God, Hope is hope for (union with) God, etc. But in any case, McPherson's objection applies only to this one particular hope, and further, only to those who hold that it is a "Theological Virtue" in the sense of an "infused" virtue. Let us therefore drop these two restrictions, and consider the general question: is there ever a duty to hope; and if so, when?

Since "*A* hopes that *P*" entails and is entailed by "*A* thinks that *P* is probable and *A* wishes that *P*," "*A* ought to hope that *P*" entails "*A* ought to think that *P* is probable and *A* ought to wish that *P*." But can these things rightly be said? It seems that the former can, but that the latter cannot; so that "*A* ought to hope that *P*" cannot be said either. Now to show this.

The "ought" in "*A* ought to think that *P* is in some degree probable" is not moral but prudential; for its force is "given the evidence that he has, *A* ought, if he is reasonable, to think that *P* is in some degree probable." It may be objected, however, that the "ought" is moral because the obligation to be reasonable (which means, principally, the obligation to proportion the degree of one's probability-estimate that *P* to the degree of probability of *P*) is a moral one. The reply to this objection is that the obligation to be reasonable is not moral but prudential, since the reason why *A* ought to proportion the degree of his probability-estimate that *P* to the degree of probability of *P* is because it is in his interest to do so. My view that the obligation to be reasonable (i.e., to proportion the degree of one's probability-estimate that *P* to the degree of probability of *P*) is prudential and not moral, is obviously similar to Price's contention that the obligation to

¹⁸ William Kneale, *Probability and Induction* (Oxford, 1949), Pt. III, sect. 34.

¹⁹ T. H. McPherson, "Christian Virtues," *Proceedings of the Aristotelian Society*, Supplementary Volume XXXVII (1963), pp. 51-68.

believe reasonably (i.e., to proportion the degree of one's belief that *P* to the degree of probability of *P*) is prudential and not moral.

Nevertheless, it may be counter-objected that the obligation to be reasonable is moral after all, because Prudence is itself a moral virtue. The Christian moral theologians, for instance, as we have just seen, list Prudence along with Courage, Temperance, and Justice as a "Cardinal" (moral) Virtue. The counter-reply to this counter-objection is as follows. The counter-objection raises the question of the relation of Morality to Prudence, which is much too large to explore now. Here, I can only give my opinion that Prudence and Morality are mutually exclusive and coordinate concepts; that Prudence is accordingly not a moral virtue; and that prudential rights and obligations cannot be reduced to moral rights and obligations, or conversely.

It is in order, then, to say that *A* ought to think that *P* has some degree of probability. But is it in order to say that he ought to wish that *P*? The difficulty, of course, is again over "ought" implying "can"; the question is whether our conscious wishes are under our control. (I insert "conscious" because nobody, presumably, would wish to maintain that our unconscious wishes are under our control.) It seems that, generally speaking, our conscious desires are not under our control either, so that it makes no sense to say that *A* ought, or ought not, to wish that *P*. Hence, it is pointless for Aquinas to say "You ought to hope for eternal happiness in the form of union with God" to a man who does not want it.

But it is necessary to add two elucidations. First, the assertion that men's conscious desires are not in general under their control means that it is not in general within their power to determine whether they have, or do not have, certain conscious desires. It does not mean, of course, that men are never capable of resisting their desires. Secondly, it is sometimes possible to attach a meaning to "you ought (or ought not) to want *X*." E.g., *A*, a frustrated militarist, might say to *B* "I wish that there would be another World War"; to which *B* might reply, "You ought not to wish for any such thing." But the sense of *B*'s reply surely is "You ought not to be the sort of person that you are, namely, the sort of person who desires another World War." And this makes sense, since *A* could, perhaps, have prevented himself from growing into a person of the sort who desires another World War.

²⁰ Bertrand Russell, *The Philosophy of Logical Atomism* (London, 1918), Pt. IV; reprinted in R. C. Marsh (ed.), *B. Russell: Logic and Knowledge* (London, 1956).

III. CONSTRUCTIVE ANALYSIS

Preliminary Terminological Clarification

We have taken the standard use of the word "hope" to be that in the formula "*A* hopes that *P*," where some person *A* is the subject of hope and some proposition *P* is the object of hope. Propositions exemplifying this formula belong to the non-truth-functional species of the genus which Bertrand Russell²⁰ calls "propositions containing more than one verb." To illustrate this point, compare the non-truth-functional species illustrated by "Tom hopes that he will recover" with the truth-functional species illustrated by "If Tom recovers then he will be glad."

There are other familiar constructions which are equivalent to "*A* hopes that *P*"; namely, "*A* hopes for *E*" (where *E* is an event-variable) and "*A* hopes to *V*" (where *V* is a deed-variable). Clearly, "Tom hopes that he will recover," "Tom hopes for his recovery," and "Tom hopes to recover" are all synonymous. Hoping-to is a special case of hoping-for, since deeds are a species of events. Whether one regards hoping-that or hoping-for as primary seems to be immaterial; that is, one can say indifferently that the object of hope is a proposition or that it is an event.

There is another use of "hope" in which "hope" is followed by a "that" clause which must be distinguished from "*A* hopes that *P*," namely, "there is hope that *P*." Here, there is no subject of hope, and the word "hope" is a noun, not a verb. In this use, the main constituent of the meaning of "hope" is "probability." E.g., the meaning of "there is very little hope that *P*" is "the truth of *P* is desirable but very improbable." Possibly the existence of this sense of "hope" explains the paradoxical-sounding locution "hoping against hope." To hope against hope is to hope against the odds or probabilities—which, as we have seen, may be a reasonable thing to do. There is an important difference between "*A* hopes that *P*" and "there is hope that *P*" in respect to objectivity. If *A* says "I have little hope that *P*" whereas *B* says "I have high hopes that *P*," they are not disagreeing with one another. But if *A* says "there is little hope that *P*" whereas *B* says "there is high hope that *P*," they are disagreeing with one another.

Hoping-that and its equivalents must be distinguished from hoping-in. E.g., Aquinas considers the question, "Is it lawful to hope in Man?" and

correctly identifies it with the question, "Is it lawful to trust in a man?". The object of hope-in is itself commonly called a hope; i.e., there are locutions of the form "*B* (or *X*) is *A*'s hope." Thus, St. Paul²¹ writes of "Lord Jesus Christ, which is our hope." Some languages, such as Greek and French, resemble English in making explicit the difference between hoping-that (or hoping-for) and hoping-in. But other languages, such as Latin and German, do not mark the difference. In German, for instance, one says both "*ich hoffe auf das Beste*" and "*ich hoffe auf Gott*."

The Meaning of "Hope"

"*A* hopes in some degree that *P*" entails (1) "*A* wishes in some degree that *P*" and (2) "*A* thinks that *P* is in some degree probable." These two tests or conditions of the truth of "*A* hopes that *P*" are severally necessary and, it is submitted, jointly sufficient. They may conveniently be called the desiderative and the estimative tests respectively. There are, however, two further tests, the evidential test and the action test.

"*A* thinks that *P* is in some degree probable" presupposes (in Strawson's sense of "presuppose") "*A* thinks that he has evidence relevant to *P*." So that if *A* says that he hopes that *P* although he thinks that he has no evidence for or against *P*, he must be disbelieved.

Again, "*A* wishes that *P*" entails "*A* is disposed to try to bring it about that *P*." So that if Tom says that he hopes to recover, but disobeys all his doctor's orders, he must be disbelieved. However, this test does not apply to all cases of "*A* hopes that *P*," but only to those where the object of hope is an action or passion of the subject; in other words, to cases of hoping-to. It applies, for instance, to "*A* hopes to get a good degree next summer" and to "*A* hopes to be operated on next week," for the reason that *A* can try to bring these events about. But it does not apply to "*A* hopes that *B* will get a good degree next summer" or to "*A* hopes that the sun will shine tomorrow," because there is nothing that *A* can do to bring these events about. Linguistically, this is reflected in the facts that one can say either "*A* hopes that he will get a good degree" or "*A* hopes to get a good degree"; but that one cannot construct a sentence using "hoping-to" corresponding to "*A* hopes that *B* will get a good degree" or to "*A* hopes that the sun will shine." Probably this is the distinction which Aquinas has in mind when, in discussing the possibility of the object of hope, he distinguishes

those objects which are producible by a human agent from those which are producible by God. For his distinction corresponds to the distinction between, on the one hand, objects of hope which are (predominantly) the subject's actions and (sometimes) the subject's passions; and, on the other hand, objects of hope which are all other sorts of events. For, since God is omnipotent, the class of events producible by God is identical with the class of all events whatever.

Questions of the form "What is *X*?" tend to be frowned on by philosophers, and rightly. Yet the question "What is Hope?" can be interpreted in a way which is illuminating, namely, as asking what sort of a thing Hope is and, by implication, is not. Taking the negative point first, I have argued against Hume and the experimental psychologists that Hope is not a sort of feeling (i.e., emotion); against Mill, that Hope is not a sort of imagining; and against Bain that hoping is not a sort of believing. (Bain's thesis will be discussed further below (in Sect. 3.4)). On the positive point, it is only necessary to say that hopes are a sort of dispositions. This means that "Mary is hoping to go to London tomorrow" does not describe a contemporaneously occurrent process in the way that "Mary is washing her hair" does, but says rather that, under certain conditions, Mary would *V*. E.g., if asked whether she wanted to go to London tomorrow, she would answer "yes." This is because "*A* wishes in some degree that *P*" and "*A* thinks that *P* has some probability" are both dispositional statements. They can be asserted truly of *A* when he is, say, asleep and is not attending to the proposition *P* at all.

Hope has four important contraries, namely, Fear, Resignation, Despair, and Desperation. The resemblances and differences between this family of concepts are best brought out by giving their respective analyses, thus: (1) "*A* hopes that *P*" means "*A* wishes that *P* and thinks that *P* is in some degree probable." (2) "*A* fears that *P*" means "*A* wishes that not-*P* but thinks that *P* is in some degree probable." (3) "*A* is resigned that *P*" means "*A* wishes that not-*P* but thinks that *P* is certain." Hence, Resignation is the upper limit of Fear. (4) "*A* despairs that *P*" means "*A* wishes that *P* but thinks that *P* is contracertain." (5) Likewise, "*A* is desperate concerning *P*" also means "*A* wishes that *P* but thinks that *P* is contracertain." Yet Desperation is not the same as Despair. The essential difference relates to action. If *A*, in a very tight corner, despairs of saving his life, he will do nothing to bring

²¹ I *Timothy*, i, 1.

about his safety, because he thinks it contraccertain that he will thereby save his life. But if *B*, in the same situation, is desperate concerning the saving of his life, he will do anything to bring about his safety, notwithstanding that he thinks it contraccertain that he will thereby save his life. In the despairing man, the estimative element predominates over the desiderative element; but in the desperate man, the desiderative element predominates over the estimative element.

The Right and Duty to Hope

A never has a duty to hope that *P*, since "ought" implies "can," and it is in general not in *A*'s power to determine whether he has or has not a certain desire. I say "in general" because it seems that it can sometimes be done. Some men can perhaps eradicate their desire to smoke—which is usually, of course, a different thing from curing themselves of the smoking habit.

On the other hand, *A* has a right to hope that *P* when: (1) *A*'s wish that *P* is neither immoral nor imprudent, and (2) *A*'s hope that *P* is reasonable. This right is always at least partly prudential, but sometimes partly moral as well. Condition (2) manifestly calls for an analysis of "reasonable hope." "*A*'s hope that *P* is reasonable" is true when: (1) *P* has some degree of probability, however small; and (2) the degree of *A*'s probability-estimate that *P* corresponds to the degree of probability of *P*; and (3) *P* is consistent both with itself and with *A*'s other objects of hope, *Q*, *R*, etc.

Conditions (2) and (3) resemble those formulated by Price for the truth of "*A*'s belief that *P* is reasonable." But this is not true of condition (1), as we shall see shortly. The interesting feature of condition (1) is that it brings out the significant fact that *A*'s hope that *P* may be reasonable even though *P* is very unlikely. Considered linguistically, the point is simply that there is no paradox, impropriety, or infelicity in saying that Dick's hope of recovery is reasonable even though the chances are heavily against it. But lying behind this linguistic fact are two non-linguistic ones. Namely, (1) the important truism that the improbable sometimes happens; and (2) the facts that, as Aquinas observes, Hope aids action by causing the subject to concentrate on overcoming the difficulty, and by virtue of its pleasant hedonic tone, which makes for more effective action. He might have added that Hope also aids action by tending to make the subject

persevere; for dogged does it. These are the two ulterior reasons why it may be reasonable to hope for the improbable. There are, however, important differences between hopes to which reason (1) applies and hopes to which reasons (2) apply. First, reasons (2) apply only to cases where the object of hope is one of the subject's own actions. Secondly, hopes to which reasons (2) apply have the interesting effect of increasing the probability of the object of hope by the very act of hoping for it. Plainly, the probability that the sun will shine tomorrow is not affected by my hoping that it will. But the probability that I shall run a mile tomorrow in less than 10 minutes does tend to be increased by the very fact that I hope to do so. In this respect, the second sort of hopes, which are self-fulfilling, resemble the "self-certifying" beliefs to which James, and more recently, C. K. Grant²² have drawn attention. Similarly, the corresponding sort of fears tend to be self-fulfilling, since Fear tends to inhibit effective action just as Hope tends to promote it. The rock-climber who fears that he will fall tends to increase thereby the probability of the very eventuality which he fears.

Hope and Belief

There are several similarities between Hope and Belief. As the standard use of "hope" is illustrated by the formula "*A* hopes that *P*," so the standard use of "belief" is illustrated by the formula "*A* believes that *P*." In the one case there are a subject and an object of hope, and in the other case a subject and an object of belief. Again, propositions of the form "*A* believes that *P*" resemble propositions of the form "*A* hopes that *P*" in being members of the non-truth-functional species of the genus propositions containing more than one verb. Yet again, beliefs, like hopes, are dispositions; and there are degrees of belief as well as of hope.

As there is both hoping-that and hoping-in, so there is also believing-that and believing-in. Indeed, the standard use of "believe-in" is synonymous with "hope-in," namely, "trust in." But this is not the only use of "believe-in," as may be seen by contrasting the statement "Joe still believes in Harold Wilson" with the statement "Joe still believes in Father Christmas."

The evidential test applies to Belief as well as to Hope. That is to say, if *A* states that he believes that *P* although he thinks that he has no evidence relevant to *P*, he must be disbelieved. For "*A*

²² C. K. Grant, "Belief and Action," Durham, 1960: H. H. Price, "Professor Grant on Belief and Action," *Durham University Journal*, vol. 53 (1961), pp. 97-202.

believes that *P*" entails "*A* thinks that *P* is more probable than not," which in turn presupposes (in Strawson's sense of presuppose) "*A* thinks that he has evidence relevant to *P*."

Again, as we have already seen, two of the tests of the reasonableness of a hope are also tests of the reasonableness of a belief. One of these is the consistency test, which requires an object of *A*'s belief, *P*, to be consistent both with itself and with other objects of *A*'s belief *Q*, *R*, etc. The other is the correspondence test, which requires the degree of *A*'s belief that *P* to correspond to the degree of probability of *P*. It follows from this correspondence test for the reasonableness of a belief that *A*'s belief that *P* is not reasonable when there is no evidence relevant to *P*; for in this case the question of the probability (or improbability) of *P* does not arise. Naturally, this applies to that special case, considered by Mill, of there being no evidence relevant to *P* in which the evidences are mutually destructive. Hence, as Anthony Flew²³ points out, it will not do to argue that it is reasonable to believe that God exists because the evidence *pro* is equal to the evidence *con*. On the contrary, what the reasonable man does in such a situation is to refrain from belief or to "suspend judgment." ("Judge," like "think," is a synonym of "believe," and is so used by, e.g., Locke.)²⁴ In this sort of situation the reasonable man also refrains, of course, from disbelief. This brings out another resemblance between Hope and Belief; namely that, as Fear is contrary but not contradictory to Hope, so Disbelief is contrary but not contradictory to Belief. Suspense of judgment about *P* where Belief is concerned corresponds to indifference about *P* where Hope is concerned.

Finally, as we have just seen, one species of hopes, namely the self-fulfilling species, in which the object is an action of the subject, are such that the probability of the object tends to be increased by the very fact that the subject hopes for it. In this respect they resemble one species of beliefs, namely the "self-certifying" species, in which the object of belief tends to be made true by the very fact that the subject believes it.

However, there are also important differences between Hope and Belief, and these must be pointed out to justify my dissent from Alexander Bain's definition of Hope as "belief in some contingent future bringing good." But first it is neces-

sary to mention, if only to dismiss, Hume's view of the distinction.

Manifestly, the essential difference between Hope and Belief on Hume's view is that Hope is an emotion whereas Belief is a sensation. Both are feelings, but in different senses of that ambiguous word. For Hume equates "*A* believes that *P*" with "*A* feels sure that *P*," and misassimilates "feeling sure" to "feeling (say) cold, or giddy." But this way of drawing the distinction between Hope and Belief will not do, since it is as false that Belief is a sensation as it is that Hope is an emotion.

The chief genuine differences between Hope and Belief are, then, these. "*A* hopes that *P*" entails "*A* wishes that *P*," but "*A* believes that *P*" does not entail "*A* wishes that *P*." Consequently, there is no desiderative test of Belief in the way that there is such a test of Hope. Naturally, one, but not the only way of applying the desiderative test is simply to ask *A* whether or not he wishes that *P*.

Since "*A* believes that *P*" does not entail "*A* wishes that *P*," it also does not entail "*A* is disposed to try to bring it about that *P*." But "*A* hopes to *V*" does entail this. So that the action test for Hope does not apply to Belief. If *A* disobeys his doctor's orders, we can infer that he does not wish, and so does not hope, to recover; but we cannot infer from the same fact that *A* does not believe that he will recover. For it may be that he believes that he will recover, but disobeys his doctor's orders because he does not want to recover. Notice that the point is that the action test for Hope does not apply to Belief. It is not denied that some other action test or tests may apply to Belief. The relation of Belief to action is, of course, a large question, and moreover, according to R. B. Braithwaite,²⁵ a complex one, since it cannot be grasped without an understanding of at least the rudiments of Decision Theory. But there is no need to go into these matters here. It is enough for the present purpose, which is to distinguish Belief from Hope, to have shown that however Belief is related to action it is not related to it in the simple way that Hope is.

The differences between Hope and Belief considered so far relate to the desiderative aspect of Hope. But we find differences no less vital when we consider Hope in its estimative aspect. "*A* hopes that *P*" entails "*A* thinks that *P* is probable in some degree, however small." But "*A* believes that *P*"

²³ Anthony Flew, *God and Philosophy* (London, 1966), p. 194.

²⁴ John Locke, *An Essay Concerning Human Understanding*, Bk. IV, ch. XIV.

²⁵ R. B. Braithwaite, "Half-belief," *Proceedings of the Aristotelian Society*, Supplementary Volume XXXVIII (1964), pp. 163-174.

does not entail this. For, although one can rightly say "*A* hopes that *P* although he thinks that *P* is improbable," one cannot rightly say "*A* believes that *P* although he thinks that *P* is improbable." On the contrary, what "*A* believes that *P*" does entail is "*A* thinks that *P* is more probable than not." Hence, Hope and Belief differ in respect to their estimative tests. Notice, in this connection, that the three following locutions are all permissible: (1) "*A* hopes and believes that *P*." This is normally said when *A* not only believes that *P*, but also wishes that *P*. (2) "*A* believes but does not hope that *P*." This is normally said when *A* believes that *P*, but does not wish that *P*. (3) "*A* hopes, but does not believe, that *P*." This is the interesting locution, since it illustrates the point that, though what the subject thinks improbable is a proper object of his hope, it is not a proper object of his belief.

Similarly, one of the tests of the reasonableness of a hope is not a test of the reasonableness of a belief. This is the test which requires that the object of hope, *P*, shall have some degree of probability, however small. For it is in order to say "*A*'s hope that *P* is reasonable although *P* is improbable." But it is not in order to say "*A*'s belief that *P* is reasonable although *P* is improbable."

Again, two of Hope's contraries, Fear and Resignation, are not contraries of Belief. For it is quite in order to say "*A* fears (or, is resigned) and believes that *P*." But the other of Hope's important contraries, Despair and Desperation, are also contraries of Belief. For one cannot rightly say "*A* despairs (or, is desperate) and believes that *P*." Notice, in this connection, that Hope and Belief are not themselves contraries; for it is quite correct to say "*A* hopes and believes that *P*."

There can be a right to believe as well as a right to hope. But whereas a right to hope is sometimes partly moral and partly prudential, a right to believe is always wholly prudential. *B*'s judgment that *A* has a right to hope that *P* is partly moral when one of *B*'s reasons for making it is his further judgment that *A*'s wish that *P* is not immoral. By contrast, "*A* has a right to believe that *P*" means "*A* is being reasonable in believing that *P*," and this right is prudential, not moral, because believing reasonably is in our interest.

Lastly, there is never a duty to hope, but there can be a duty to believe. The duty to believe, like the right to believe, is purely prudential. For "*A* ought to believe that *P*" means "*A* ought, if he is reason-

able, to believe that *P*"; and the obligation to be reasonable is prudential, not moral, because believing reasonably is in our interest. But since "ought" implies "can" and "*A* hopes that *P*" entails "*A* wishes that *P*," there can be no duty to hope because, generally speaking, it is not in one's power to determine whether he shall or shall not have a certain desire.

It remains to direct attention to one further, and highly important, connection between Hope and Belief, namely, that Belief forms part of the meaning of Hope. For a hope includes a probability-estimate, and a probability-estimate is precisely a believed-probability. In other words, "*A* hopes that *P*" entails "*A* thinks (i.e., believes) that *P* has some degree of probability, however small." As we saw when criticizing Hume's philosophical analysis of Hope, Hope cannot be defined in terms of only the two concepts, Desire and Probability, as he supposed. But it can be defined in terms of the three concepts, Desire, Probability, and Belief, above.

IV. CONCLUSION

We have seen that, whereas it is not correct to say that *A*'s belief that *P* is reasonable although *P* is improbable, it is correct to say that *A*'s hope that *P* is reasonable although *P* is improbable. This fact lends considerable interest to Kant's and Mill's opinions about religious hopes. Venturing a crude sketch of a complex subject, I suggest that it may be represented as follows. At one time it was thought that, to justify such articles of religion as the existence of God or the immortality of the soul, it is necessary to show that men have a right to claim to *know* them, since they are certain, or true, or even analytic. But men soon saw that this will not do, and that Descartes,²⁶ for instance, fails signally in his Ontological Argument to show that the proposition "God exists" is analytic and so may be claimed to be known. With Hume,²⁷ however, came a fresh view of the question. It was now held that, to justify the articles of religion, it suffices to show that men have a right to *believe* them, because they are probable in the colloquial sense of "more probable than not." Yet it was soon felt that this will not do either, since the Teleological Argument, e.g., as presented by Hume and Mill, does not in fact make it more inductively probable than not that there exists a god in the sense of superhumanly intelligent,

²⁶ René Descartes, *Meditations on First Philosophy*, No. V.

²⁷ David Hume, *Dialogues Concerning Natural Religion*, Pt. II.

powerful and (perhaps) benevolent artificer. Hume makes this quite clear.

Is it so certain, however, that this argument does not probablify that hypothesis in any degree at all, however small? Those who are interested in this question might find it to their advantage to abandon the exposed and probably untenable position that there is a right to believe the articles of religion, and to fall back on the position which has been previously

prepared by Kant and Mill, and which looks more easily defensible. This is the position that those who so wish have a right, which is partly moral and partly prudential, to cherish their supernatural *hopes*, since these are neither immoral nor imprudent, and are moreover reasonable in that sense of the word "reasonable" which applies to hopes; the sense, namely, of possessing at least some degree of probability.²⁸

Keele University

Received March 15, 1968

²⁸ This paper was read to the Cambridge University Moral Sciences Club on February 23, 1967.

II. OBJECTS AND QUALITIES

HAIG KHATCHADOURIAN

INTRODUCTION

THE ordinary concept of an object¹ is a knot that ties together the ordinary concepts of quality, property, causality, potency, power, and space and time. All these concepts find an important meeting point in the concept of an object; though (to a lesser extent) some or all of them also find a meeting point in the concepts of event, occurrence, and happening. This is not surprising, since objects are, in a basic sense, the "building blocks" of the physical world; and the concepts of quality, property, causality, potency, power, and space and time themselves have some reference or other to the concept of an object. In the present paper I shall mainly consider, as much as space permits, some aspects of the logical grammar of "object," "quality," "property," and some of their major philosophical implications. Consequently—though very briefly—I shall also consider the ordinary concepts of capacity, potentiality, potency, causality, space and time insofar as they involve a reference to objects. These concepts, which are of great importance in their own right, will be considered in some detail in subsequent papers. Likewise the logical grammar of "object," "quality," and "property," including the main differences between qualities and properties, will be further considered in future papers.

The present paper is partly a logical continuation of and sequel to my analysis in "Relations"² and "Do Ordinary Spatial and Temporal Expressions Designate Relations?"³ I shall begin with a brief review of some of the results arrived at in the latter paper which conveniently serve to launch the present inquiry.

I

In the paper alluded to I attempted to show,

among other things, that in the ordinary uses of "relation" the spatial or quantitative *parts* of a physical object, say a table, are not *related* to one another (and to the object as a whole)⁴ by the mere fact that they are spatially continuous with or connected to one another, and co-exist in time. This is one spatial application of the general thesis, elaborated in that paper, that spatial and temporal expressions⁵ do not designate relations in any relevant ordinary use or sense of "relation." However, the parts of an object can perhaps be empirically related to one another, and to the object as a whole, by certain causal relations.

In this inquiry I am not concerned as such with the way we ordinarily talk about the parts of an object, or with an object insofar as it has spatial parts. I am, rather, concerned in the first place with whether or not the different qualities of an object are ordinarily said to be related to one another, and to the object as a whole, *qua* qualities of the (same) object. In other words, I am concerned with whether (i) "*a* is a quality of object *O*" designates some sort of relation between *a* and *O*; and with whether (ii) "*a* and *b* are qualities of *O*" has reference to some sort of relation between *a* and *b* themselves. The same questions can be phrased in terms of *O vis-à-vis* its qualities; viz., whether "*O* has (or possesses) qualities *a*, *b*, *c*, *d*, etc." refers to (*A*) some relation between *O* on the one hand and *a*, *b*, *c*, *d*, etc., on the other; and whether it refers to some relation(s) (*B*) between *a*, *b*, *c*, *d*, etc., themselves.

My view is that the answer to both of these questions is the same, *mutatis mutandis*, as the answer I had earlier given with regard to the parts of an object and the object as a whole; namely, that these expressions do not refer to any (empirical) relation(s) in any ordinary sense or senses of "relation." That is, the qualities of an object do not, simply *qua* qualities of an object, stand in any empirical relation

¹ I generally use the word in an inclusive sense, to designate (among other things) both what we ordinarily call objects (solids and some semi-solids) and stuffs and substances (liquids and some semi-solids); though I use it in the more usual way whenever necessary. I also distinguish (in Part II) objects and substances in the chemist's technical use of "a substance" and "substance."

² *The Southern Journal of Philosophy*, vol. 19 (1964), pp. 133-142.

³ Unpublished.

⁴ Though they are either spatially *continuous* with or *connected* to one another. But neither—not even the latter—constitutes or involves a *relation* between the continuous or the connected parts, *qua* continuous with or connected to one another.

⁵ Leaving aside the adjectival uses of number-name (e.g., "69" in "69 years"), which I did not consider.

to the object that has them. Similarly, the qualities of an object are not related to one another, *qua* qualities of the same object. However, some and perhaps all of these qualities may be causally related to one another, in some sense or other. I shall return to this point a little later.

The reason for my view is simply that, as far as I can see, we do *not* ordinarily speak of or regard qualities as related to one another or to an object *O* insofar as they are qualities of *O*; "*a* is related to *O* as a quality of *O*" and "*a* and *b* are related to each other as qualities of *O*," I submit, have no meaning, or are improper, in ordinary language. If we divest ourselves of the way traditional philosophers have tended to talk and think about relations,⁶ the unusualness of this way of talking becomes, I think, apparent.

(1) The foregoing contradicts what has been held by many classical philosophers about objects and their qualities. For instance, this is true with regard to the various forms of substantialism⁷ and of insubstantialism in the history of philosophy. Both substantialism (e.g., in Descartes and Locke), and insubstantialism (e.g., in Hume) *commit the same basic error*. Both of these types of theory suppose that we can properly speak of the qualities of an object as related to one another (and for the substantialist, to the alleged material substratum). This supposition⁸ constitutes one main logical ground for the substantialist's view that there must be a material substance to hold together or unite the qualities of an object, to make them the qualities of *an* object. These qualities are supposed to be *related* to the material substratum as predicates inhering in a subject.⁹ By virtue of this, if for no other reason, they are also supposed to be *related* to one another, as the qualities

of the same object. But once it is seen that this is an improper way of talking or conceiving of qualities and of objects, the question no longer arises as to what makes possible the relations among the qualities of an object. Consequently, one of the main reasons for positing a material or a mental substratum is eliminated. The "problem" as to what gives unity to the qualities of an object—to which the Cartesian-Lockean conception of a material substance and Berkeley's conception of a mental substance¹⁰ were regarded as the correct solution—is seen to be spurious. The problem will then tend to "dissolve," and the temptation to posit material or mental substance will be more easily overcome. Stated positively and more directly, what I am saying is that the ordinary uses of "object" do not commit us to substantialism (or to insubstantialism).

(2) Berkeley, who rejects the Cartesian-Lockean material substance, commits the error no less than Descartes and Locke. He too supposes that (a) the qualities and properties of what we ordinarily call an (a perceived) object are *related* to one another as the qualities and properties of an object, and (b) attempts to meet this in what is logically a fundamentally similar way. He too posits a substance as the substratum of these qualities and properties, of which they are some of the modifications. The basic difference between him and Descartes or Locke is that he posits *mental* substance as the substratum of these qualities and properties; i.e., regards a sensible object as a cluster of "ideas" existing in the mind. We are not here concerned to ascertain the correctness or incorrectness of his view that perceived qualities, and so (for him) perceived objects, are mental. Nor are we concerned to ascertain whether or not there is such a thing as mental sub-

⁶ I am of course suggesting by implication that traditional philosophers have quite drastically modified (broadened) the ordinary uses of "relation" by assimilating to relations the uses of certain expressions that do not designate relations. This is what I maintained concerning "similar" and "resembles," and "different from," etc., in my "Relations." Similarly with spatial and temporal expressions, in my "Do Ordinary Spatial and Temporal Expressions Designate Relations?"

⁷ I use "substantialism" here in a very broad sense, to designate the view which holds that objects consist of some substance—*whether material or mental*—in which qualities inhere. Thus Berkeley's view of *perceived* objects is a form of substantialism in this sense, as much as Descartes' and Locke's view of "*material* objects." I must also explain that in this paper I use "object" to refer to perceived objects, following ordinary usage. But whenever the epistemological dualist's distinction between perceived (or sensible) objects and "material objects" is relevant, I shall qualify "object" as required.

⁸ While one of the main reasons for this supposition itself is the general assumption, shared by the classical substantialists and insubstantialists alike, that sensible qualities are *distinct* or "*atomic*" sense impressions *separately* "received" by the mind through one or more of the senses.

⁹ For example, Locke, *Essay Concerning Human Understanding: From Descartes to Locke* (Chicago, 1960), ed. by T. V. Smith and Marjorie Grene, pp. 356, 359, 374, 381–382, etc.; Berkeley, *First and Third Dialogues Between Hylas and Philonous: Berkeley, Hume, and Kant* (Chicago, 1960), ed. by T. V. Smith and Marjorie Grene, pp. 28–29, 32–34, 73, 81, etc.

Note also Aristotle's conception of the alleged relation between form and matter with regard to any substance, including objects; his conception of the category of substance and the predicables; and the conception of such singular subject-predicate propositions as "The table is brown" in traditional Aristotelian logic.

¹⁰ Actually, as is commonplace, Descartes' and Locke's *mental* substance, like Berkeley's mental substance, is partly intended to account for the existence and alleged unity—as well as for the alleged ideal nature—of the *perceived* qualities of objects.

stance alleged to be the "subject" of these "predicates." But to the extent to which Berkeley's conception of objects rests on the supposition that the qualities of an object are related to one another, and so to the extent to which his view that they inhere in some substance (mental substance) rests on this supposition, it is logically confused and uncalled for.

(3) It may be urged that Hume certainly did not commit the error in question, since he expressly rejects the view that an object is more than a bundle of sensible qualities (which he, like Berkeley, regards as mental). That is, he regards an object as a bundle of mental "impressions." He rejects Berkeley's view that mental substance exists, since (among other things) he holds that sense-impressions (sensible qualities) are *distinct from one another, and can and do exist apart from one another*. Hence he sees, in effect, no conceptual need for positing a mental (or a material) substratum to relate them to one another as impressions of what we ordinarily call the same object.¹¹

This is all true, and to this extent he escapes the error alluded to; but he does not escape it altogether. For (a) in his rejection of Berkeley's view, and (b) in his positive view, he too assumes that the problem which gives rise to Berkeley's (and, in a different way, to Descartes' and Locke's) view about substratum is a genuine problem. What he therefore does is to give us what he believes is a radically different *solution* to it. For though he rejects the view that the qualities of an object are related to one another as modifications or predicates of some substance, he does think that they are related to one another in certain ways; namely, by spatial contiguity, resemblance, and causality. For he regards both contiguity and resemblance (as well as causality, which is a genuine relation) as relations.¹² This does not give an object the kind of unity which the substantialists claim it has; though he thinks that it does give it a unity of sorts. But he is also anxious to show how we form the *idea* of an object as something self-

identical through time; and so he gives us his psychological, associationist theory.

According to this theory, as will be remembered, the unity we ascribe to objects is only an idea of the imagination. The idea of numerical identity results from the operation of the principles of association with the indispensable help of memory. That Hume himself was eventually dissatisfied with this account¹³ is significant. But what is more noteworthy here is that his associationist theory itself is his *main positive answer* to what he thought was a genuine problem; namely, accounting for the apparent unity of objects. The substantialist's type of account attempts to explain two related things in one stroke: (1) the alleged relatedness of the qualities of an object, and hence (2) the apparent self-identity of many objects (certain conditions being fulfilled¹⁴) through time.¹⁵

The apparent identity of an object that we ordinarily describe as the same during some stretch of time, is, on the substantialist type of view, a *consequence* of the alleged relatedness to one another, with the help of the common substratum, of the different qualities of the object. The latter remains the same because its substratum remains unchanged; though some or all of its specific qualities may suffer some change.

(4) Kant's conception of perceived objects is also based in part on the mistaken supposition that the qualities of an object, as a perceived object, are related to one another in some way, forming the unity which an object consists in.¹⁶ He also accepts the general view of Descartes, Hobbes, Locke, Berkeley, and Hume that the manifold or "given" element in perception (which he regards as the matter of that which is perceived) is a set of discrete, "atomic" sense impressions. But he rejects the view that the unity which they (are perceived to) have is given to them by their alleged inherence in some substratum; and at the same time he rejects Hume's view that a perceived object is nothing but a bundle of numerically distinct sense impressions (qualities),

¹¹ We should remember that he also believes that there is direct *empirical* evidence against the belief in a mental substance in Berkeley's, Descartes', or Locke's sense(s); namely, that introspection does not disclose the alleged simple continuous and unchanging self or mind.

¹² *A Treatise Of Human Nature*, Pt. IV, sect. VI, "Of Personal Identity," *passim*. Of course he then resolves causality itself into regular sequence of similar impressions.

¹³ *Ibid.*, Appendix.

¹⁴ That is, if they are not suffering such qualitative changes as would result in their becoming, as we say, something else.

¹⁵ Looking at the matter from a "linguistic," conceptual standpoint, we can say that the substantialist's and the insubstantialist's lack of awareness of the impropriety of the question: How are the qualities of an object related to one another?—which they implicitly or explicitly identified with the question: What makes an object *an* object?—logically led them in effect to change, in their different ways, the ordinary uses of "object" and "quality."

¹⁶ However, the *conception* of an object as something which consists of a subject possessing certain predicates (attributes) is the work of the understanding.

lacking all *substantial* unity. He holds that the unity which (he believes) perceived objects possess is contributed by the perceiving subject himself, though in a different way from Descartes, Locke, or Berkeley on the one hand, or Hume on the other; since he holds that the "given" is chaotic, amorphous. This unity or organization (the element of form) consists of the forms of space and time, the two *a priori* forms of sensibility. The ordering of the sensible qualities of an object, and of objects as a whole in relation to one another, is the work of these ideal forms. The mind organizes what is given by relating the elements to one another spatially and temporally.¹⁷ The process of synthesis is continued further by the application of concepts¹⁸ (and hence the categories of the understanding). Thus to apply the concept of an object to something is to conceive it as a synthetic unity, a unity of matter and form.

(5) A careful consideration of sense datum theories is beyond the scope of the present discussion. What concerns us here is just one aspect of these theories, which I think is common to them all; namely, that the sense data of an object *O* are somehow *related* to *O* in the case of "veridical," e.g., non-illusory and non-hallucinatory, perception. In cases of "non-veridical" perception, there is no such relation because the sense data in question will not be the sense data of any object or any other kind of existing thing. The kind of relation that is thought to obtain in "veridical" perception varies with individual theories; and in the case of G. E. Moore it is well known that he tended to oscillate between two different views of the nature of this relation. Further, as Ryle points out, "Disputes continue about the linkages obtaining between [say] horse-races, which [according to this theory] we do not 'strictly' or 'directly' see, and the looks of them [the sense data] which we do 'strictly' or 'directly' see, but which are not race-courses."¹⁹ But insofar as all of these theories envisage some relation between sense data, and, at least in cases of "veridical" perception, between them and the objects of which they are the sense data, the basic criticism I made earlier with regard to the relevant views of Descartes, Locke,

Berkeley, Hume, and Kant also applies here.²⁰ I should add that this criticism would become superfluous if it is shown that (as e.g., Ryle and Austin, among others, attempt to do) talk about sense data is misguided, or if—to state it in more traditional terms—sense data do not exist.

II

Part I has dealt with a basic error commonly committed by philosophers, generally as a result of their misconstruing the uses of "quality" as well as of "relation" in their everyday employment. In this section I shall do two related things. I shall first consider explicitly some of the main differences between the uses of "object" and "quality." In this way the uses of both expressions will be brought into relief; and this will help to show rather dramatically the extent to which insubstantialism and phenomenalism *modify* ordinary language. Or stated in terms of the ordinary uses of "object," the foregoing will show that an object cannot be exhaustively described purely in terms of the kinds of qualities it has. What the concept of an object includes over and above the concept of "something that has certain kinds of qualities, *a, b, c, d*, etc.," will be discussed below (in sect. 2 of Pt. II).

1. Qualities versus Objects

(1) The first thing to note here about the logical grammar of "quality" is that we do not ordinarily speak of the qualities of an object as *coexisting in the same region of space*. We ordinarily say, of course, that a table, say, is at the same time both red and round; but we do not ordinarily say that the redness and the roundness of the table themselves simultaneously exist. If we do say this, it would be merely (I think) an awkward way of saying that the table is both red and round. The redness and the roundness of a table are *not things that can be strictly or literally said to coexist*. (See later.)

Again, the concept of a place, and so the concepts of "same place" and "different places," apply to

¹⁷ Cf. "perceptual judgments 'merely express a relation of two impressions to the same subject . . .'" S. Körner, *Kant* (London, 1955), p. 53.

¹⁸ The imagination first "collects" the sense-presentations, which are then, logically speaking, given their conceptual unity.

¹⁹ Gilbert Ryle, *The Concept of Mind*, ch. 7, quoted from *Readings In Twentieth Century Philosophy* (New York, 1963), ed. by W. P. Alston and G. Nakhnikian, p. 592.

²⁰ The sense datum theory, in the present and in current philosophical discussion, should be distinguished from the theories of Berkeley and Hume. The latter, to the extent to which they can be cast in linguistic terms—as a recommendation for an alternative language of perception—would now be called a version of phenomenalism (though, in the case of Berkeley, only by leaving out his view that sensible qualities inhere in mental substance).

objects but not to qualities (or properties) *as such*.²¹ We speak of a red object, say, as being in some place at a given moment of time; but we do not ordinarily say that the redness of the object is in some—in that—*place*. However, we do say that a certain *patch* of color, a colored *spot*, a *daub* of color or a *stain* is to the left, right, etc., of another patch of color, spot, daub, stain, or to the left, etc., of an object; just as we say that an object is to the left, right, etc., of another object. Further, we speak of the relative spatial positions of the *lines*, the *masses* of color, the different colored *forms* in a painting (which have some color or other). Similarly with the colored spots on a polka-dot dress, the colored stripes of a tiger or zebra, the colored spots of a leopard, the colored plumes of a bird; and so on.

On the other hand, I do not think we speak of the color(s) of an object as *occupying* the *same place* as the *colored surface*; or of the color of an object as occupying the same place as its shape, size, hardness, smoothness, and so on. Nor do we speak of these qualities themselves as occupying some place or other at a given time.

(2) The notions of existence and nonexistence are not ordinarily applied to the qualities of objects, but only to objects as a whole. Similarly, these notions are not ordinarily applied to properties, but only to the things that have them. Philosophers, of course, speak of certain qualities and properties as existing or not existing. But what meaning does "Quality *Q* exists (does not exist)" have in *ordinary* discourse? It seems to me that when—or if—we speak in this way in everyday contexts, this would be an awkward and roundabout way of saying that there is some existent object, person, occurrence, and the like which has quality *Q*. When we say that the quality "red" exists, while the property "being square and circular" does not exist, I think we ordinarily mean that there are red things (red things exist), but that there are no square circles (no square circles exist).²²

These things apply to qualities and properties as "particulars," as "instances" of redness, hardness, triangularity, etc. The situation would be different if, with some philosophers, we conceive of qualities and properties, *generically speaking*, as Platonic universals. For on this view it is true that—hence it is

meaningful to speak of—redness, hardness, etc., as existing (or rather, as "subsisting"), as having a reality of their own, apart from their particular "instances" or "exemplifications," and so apart from all existing objects, persons, occurrences, and the like.

Whatever the merits or demerits of Platonic realism, what we said about the ordinary employment or lack of employment of "exists" in relation to qualities and properties remains. For our remarks concern qualities as "particulars"; specifically, the qualities of individual objects, *qua qualities of the latter*. Similarly, *mutatis mutandis*, with properties. In other words, they concern the question of whether, say, we speak of the redness of the lampshade before me as existing—and if so, what we mean by this—as opposed to speaking of the lampshade (as a whole) as existing.

A. J. Ayer writes the following regarding the question of the existence of the properties: "It might be thought that we could discriminate individuals from properties by saying that only individuals existed . . . The suggestion must be that it always makes sense to say of an individual that it exists but never makes sense to say this of a property."²³ Ayer states two objections to this. The first is ". . . that it is still a matter of dispute whether all the things that are commonly reckoned as individuals can significantly be said to exist."²⁴ Examples are spatio-temporal points, biological species, and unconscious minds. This objection clearly does not touch *objects*, in the ordinary or in our extended use of the word, in whose case it always makes sense to say that they exist (or do not exist). The second objection is that ". . . it is by no means common ground that properties do not exist. A great many philosophers have held that they do, and I do not see how they can very well be refuted."²⁵ This objection, for what it is worth, is irrelevant to the question of whether we ordinarily speak of properties (and qualities), *qua particulars*, as existing or not existing; since it pertains to properties as (alleged) universals. Ayer does not distinguish the question of whether it ordinarily makes sense to speak of properties, *qua* (alleged) universals, as existing or not existing, and the parallel question regarding properties *qua particulars*.

²¹ But in the case of sunsets and rainbows, which are not objects in the ordinary meaning of "object," we refer to the relative spatial positions of the colors of a sunset or a rainbow, e.g., in trying to point out some of their features to an onlooker.

²² We do ordinarily apply the notion of existence to sounds; e.g., in "There is a buzzing sound in the basement." But sounds are not ordinarily regarded as qualities or properties of objects, strictly speaking; though the "capacity" to emit sounds is regarded as a property of objects. On the other hand, smells, heat, and cold do not constitute exceptions to what I said above.

²³ "Individuals," *Philosophical Essays* (London, 1963), p. 3.

²⁴ *Ibid.*

²⁵ *Ibid.*, p. 4.

The discussion of the existence or nonexistence of properties as universals does not, either directly or indirectly, answer or help answer the latter question.

Our remarks about the uses of "exists," if true, have important implications. Here I shall mention only three.

(a) The reconstruction of the ordinary concept of an object by insubstantialism and by the sense datum theory necessitates the modification of the ordinary uses of "exists," in line with the view that objects are bundles of qualities or that sense data exist, respectively.

(b) The second implication is even more important. The insubstantialist frequently argues as follows: we can conceive or imagine *all* the qualities of an object as not existing (we can imaginatively "strip" the object of all its qualities). But if this is done, nothing whatever remains; *since we perceive nothing but sensible qualities* when we perceive what we call an object. On the other hand, if our above statements about "exists" are true, to conceive or imagine *all* the qualities of an object as not existing is necessarily nothing more or less than to conceive or imagine the object which has these qualities as not existing. The insubstantialist supposes that to conceive or imagine all the qualities of an object as not existing is a distinct conceptual operation from conceiving or imagining the object as not existing; since he thinks that "All the qualities of an object *O* are conceived as not existing" has a distinct meaning from "Object *O* is conceived as not existing."²⁶ Consequently, he thinks that it is an empirical or contingent fact that if all the qualities of *O* did not exist, *O* would not exist. And he thinks this to be the case because, for one thing, he thinks that sense perception supports his "view" that we perceive nothing but sensible qualities when we perceive what we ordinarily call an object. As a result, he mistakenly thinks that the fact that no object remains if we conceptually "strip" an object of all its qualities *empirically shows* that an object is nothing but the sum total of its qualities in certain relations.

The substantialist likewise thinks that the conceptual operation in question consists of something other than conceiving an object as not existing. But unlike the insubstantialist, he imagines or conceives an unknown somewhat²⁷ (since he *accepts* the insubstantialist's thesis that we sensibly perceive only qualities) as remaining when the operation is

performed. This "somewhat" is partly posited to account causally for our perceptions, as well as (as I remarked before) to account for the unity and self-identity of perceived objects.

But further, the idea that nothing remains if we "strip" a *perceived* object of its qualities will not do on Descartes' or Locke's own view that extension is the essence of *material* substance. I shall only mention the one reason that is relevant to us here; namely, that if extension is the essence of material things, i.e., if it is not merely contingently but necessarily inseparable from matter, it is logically impossible to "strip" a material thing of its extension. Or, stated somewhat differently, if we (*per impossibile*) can "conceive" this process of "stripping" a material object of its extension, then, by the definition of matter, nothing would remain—but not for the reason Berkeley, for example, thinks. Nothing would remain not because there is no such thing (or there is no empirical evidence for the existence of such a thing) as material substance, but because matter cannot exist without extension, its essence.

Moreover, since shape and size, motion and rest are entailed by extension, it also follows that a material thing cannot be consistently conceived as stripped of any of these qualities.

It appears, therefore, that Locke made an undue concession in stating that matter is an unknown somewhat; since the primary qualities of a material object, together with the various "powers" it has by virtue of its primary qualities, are all the qualities a material object has; and if these are (*per impossibile*) "stripped" off, we should obviously get an unknown and unknowable somewhat. The same would be true of any other kind of existing thing, if we (*per impossibile*) "strip" it of all its defining qualities or properties.

(c) Finally, what we said has an important implication for the celebrated dispute as to whether existence is a property. For if our above thesis is or were *false*, one could readily construct a *reductio ad absurdum* argument proving that existence *cannot* be a property. But I shall not go into this argument here.

The foregoing differences and others I shall presently mention, between the logical grammar of (a) "quality" and so names of qualities (also, "property" and names of properties), and (b) "object" and so object names, show how drastically

²⁶ In other words, the insubstantialist thinks that his "view" is an empirical theory about objects (and their qualities), rather than an analysis of what we (ordinarily) mean by "object." This is the fundamental reason why the insubstantialist would resist the claim that all that he is doing is offering an alternative language, a revision of ordinary language.

²⁷ This applies to Locke but not Descartes. The difference is due to their divergent epistemological views.

ordinary language is modified by the "theory" that an object is a bundle of qualities or is partly or wholly constituted by sense data. This is nowadays most commonly expressed by saying that, in the ordinary uses of these expressions, we cannot possibly translate statements about objects into statements about qualities. Or stated in more traditional terms, that objects cannot be regarded as bundles of qualities or as partly or wholly constituted by sense data. The insubstantialist or phenomenalist, in effect, applies to qualities notions *borrowed* from those ordinarily applied to objects. In some cases this borrowing results in a clearly figurative way of speaking. The use of "bundle" in speaking of an object as a bundle of qualities is a striking example. An object is imagined as some sort of collection of *things* (objects) supposed to be somehow related; i.e., a concept applicable to objects is applied to qualities—something to which, given the ordinary literal meaning of "bundle," it does not apply.

2. *The Qualities of Material Objects*

(1) In his "Matter and Space," Antony Quinton lists the following qualities as the spatial qualities of material objects in the inclusive sense of "object," i.e., in the sense in which it refers to stuffs and substances as well as objects in the usual sense: (A) occupation of space "in the full inclusive sense" in which only things possessing volume (i.e., objects as opposed to points and surfaces) are occupants of space. He adds (in agreement with Descartes) that "To be an occupant of space and to be material are one and the same." Also, "Whatever occupies space has a shape and size." (B) "... Whatever occupies space has at any given instant a position in space, which will be a volume rather than a point, and it stands in some relation to all the other occupants of space. In virtue of this relation to other occupants of space as it persists or alters through time the material thing is at rest or in motion." (C) "Finally all the occupants of space are solid in a sense that is distinct from mere voluminousness. This type of solidity may be called impenetrability . . . It is the property that every material thing possesses of excluding every

other material thing from simultaneous occupancy of the region of space where it is to be found. The concept of logical solidity or impenetrability is implicitly defined by the principle that no two things can be at the same place at the same time unless one is a part of the other."²⁸ This group of characteristics is Locke's list of primary qualities (leaving out number), and is, for Quinton as for Locke, "definitive of a material thing, for the possession of them is entailed by the occupancy of space which in turn is the necessary and sufficient condition of material thinghood."²⁹

With a few reservations or qualifications, I am in agreement with Quinton and Locke³⁰ regarding the foregoing. One of these reservations—the one regarding the alleged *spatial relations* between material things—I spoke about in Part I. As for Quinton's view that the different parts of a thing can be (and are) simultaneously at the *same place*, this will be critically considered later on.

(2) We have just noted the "defining" (kinds of) qualities of objects in the inclusive sense of "object." We do not need to remind ourselves that many objects also possess color, odor, or taste, and emit sound under certain conditions. Earlier, in Part I, we had noted a major logical source of the erroneous view that an object consists of a material substance in which its qualities inhere. Analysis of the *ordinary* uses of "object" also shows that this word does not designate some substance in the common philosophical (e.g., the Cartesian-Lockean) sense, or any other kind of entity, distinct from the qualities it has, but which in some mysterious fashion "underlies" them. In this respect the substantialist's view, if regarded as a *description of what we ordinarily designate by "object,"* gives us, in Ayer's words, a "metaphysical invention."³¹ However, substantialism can be more appropriately regarded as a reconstruction of the ordinary concept of an object.³² On the other hand, as I also pointed out, an object is not "merely" a "bundle" of qualities. So the question now is: In what ways is the ordinary concept of an object not exhausted by the concepts of its qualities logically considered together? Or stated otherwise, what does "This is an object" logically imply in addition to "It

²⁸ *Mind*, vol. 73 (1964), pp. 341–342.

²⁹ *Ibid.*, p. 342.

³⁰ Locke, however, regards primary qualities as "definitive" of material objects in a sense of the phrase in which a material object is distinguished from and opposed to a perceived object. This is at least true in the greater part of his *Essay Concerning Human Understanding*, where he subscribes to epistemological dualism. Whereas I use "material (or physical) object" and "perceived object" interchangeably, in accordance with ordinary usage.

³¹ "The Identity of Indiscernibles," *Philosophical Essays* (London, 1963), p. 31.

³² Cf. *The Structure of Metaphysics* (London, 1963), ch. VII. This chapter, though vitiated by defective exposition and an imperfect understanding of traditional substantialist theories, contains valuable criticisms of the view.

235080

occupies space, has some shape and size, is at rest or in motion, is logically impenetrable, and perhaps has a color, an odor, a taste, etc."³³ The answer, to my mind, is the following. "X is an object" logically implies (A) that X is a relatively "stable" or "enduring" thing, using "stable" and "enduring" in a somewhat technical sense; and (B) that X has causal efficacy. I shall now explain briefly what I mean by these things.

(A) In speaking of an object as something relatively "stable" or "enduring," I mean that it makes no sense—in contrast to events, occurrences, happenings, processes, and activities—to say such things as: "This (object) started at 10 a.m. yesterday and ended at 11 p.m. this morning"; though we say: "This table was made on July 15, 1966 and was chopped to pieces on June 4, 1967." Objects come into existence; and some do so by being produced. They do not happen, though they undergo change, or all sorts of things happen to them (though events do not happen to them; nor do processes occur in them but in the substance(s) they are made of, in the chemist's sense of "substance").

(B) All subjects can causally act and be acted on by other objects. This is what Locke means by saying that objects possess "active" and "passive" powers respectively. But Locke attributes "powers" to what he calls material objects as opposed to perceived objects; whereas this notion is part of the ordinary concept of a perceived object, and there is nothing in the everyday world-view which corresponds to Locke's material objects. In a different way from Locke, and for different reasons, Hume eliminates the notion of "powers" from the concept of a perceived object. He does this by reducing perceived objects to bundles of qualities ("impressions") and, concomitantly, by reducing causality to regular sequence. At the same time, these moves are in line with Hume's (also Locke's) view that perceived objects are mental.

It should be noted that we do not ordinarily say that *all* objects have certain powers, capacities, or potentialities. These expressions are used in various restricted ways, in relation to different kinds of objects or other kinds of things, in marked contrast to their very general employment by philosophers since Aristotle in relation to all objects. In ordinary discourse we speak of the capacities or potentialities

of men and animals, especially men; and we speak of exploiting the potentialities of a situation, the economic potentialities of a country, etc. But we do not speak of the capacities or potentialities of objects; and so on.

The question now is whether we ordinarily speak of the qualities of an object or the properties of the substance(s) of which it is made as having "causal efficacy" as qualities or properties.

Now we certainly say that the colors of a flower, a landscape, or a painting; the odor of certain objects; the sound of a siren or a whistle have various psychological effects on *people*. The effect is sometimes produced by the sheer sensuous quality of the colors or odors, and not always as the colors or odors of some particular object. Similarly with sounds. Further, *some* "secondary" qualities are ordinarily said to have certain causal effects on *objects*. Thus we say that the dark color of the roof causes the attic to overheat in the summer and we say that the heat of a glass container, under certain conditions, makes the container crack. Similarly with sounds.³⁴ But we do not, I think, speak of the taste or smell of an object as having any possible effect on other objects.

The situation is more straightforward with respect to "primary" qualities. All such qualities can have causal effects on objects and persons. (With regard to the latter, what I said about color, taste, odor, and sound applies to them.) We say: "The sharpness of this knife is responsible for the gash in my hand," and "The hardness of the object A threw at the car caused the dent in the fender; and because of the object's size, the dent it made was big." And so on. But note that the sharpness of *the knife* produced the gash, and the hardness and size of *the object* caused the large dent. Of course, the sharpness of a knife's blade presupposes its other "primary" qualities, e.g., hardness, occupancy of space, etc., hence the knife as a whole as an extended object.

I said earlier that the concept of causal efficacy is an integral part of the ordinary concept of an object. Part of Hume's modification of the ordinary concept of an object therefore lay in his elimination of this notion from the concept of an object, partly through his (in effect) reconstruction of the everyday concept of causality. This, among other factors, resulted in his insubstantialism. The two aspects of Hume's philosophy go hand in hand.

³³ For a somewhat different though not altogether unrelated approach to the analysis of what is meant by an individual (object), which likewise rejects both substantialism and insubstantialism as purported descriptions of what is ordinarily meant by an individual, see A. J. Ayer, "Individuals," *op. cit.*, pp. 1-25. In his earlier "The Identity of Indiscernibles," *op. cit.*, however, Ayer is inclined to accept the insubstantialist position.

³⁴ I owe this point to my colleague, Prof. Mendel Cohen of the University of Wisconsin, Milwaukee.

III

1. *Introductory*

Although both substantialists and insubstantialists are wrong in their formulation of it, there is a genuine question which, in their mistaken way, these philosophers attempt to answer; namely, what constitutes the "unity" or oneness of an object, which has different (and different sorts of) qualities. This question is, indeed, not a single question but a cluster of related questions. For example, one of the questions involved is: What makes certain qualities *a, b, c, d, e, f*, that we simultaneously perceive in what we might vaguely call the same region of space, qualities of a single object *O*? Or why do we "attribute" all these qualities to the same object *O*? A closely related question is this: (1) What makes a particular object an individual object, numerically distinct from other objects? To answer the latter question, we must find out (2) how in fact we distinguish objects, what criteria we actually use to distinguish them, or how we identify an object as an object. For the "features" of an object that make it an individual object as opposed to a collection of objects are precisely those which are theoretically available to us for distinguishing the object from other objects, and so for correctly identifying it as an object. And some or all—usually some—of these "features," though not necessarily always the same ones, are those we actually utilize in judging that a particular object *O* is a single object.

The identification of an object as an *individual object* is logically distinct from its identification (A) simply as an *object* rather than an event, a situation, a person, and so on, and from its identification (B) as an individual table, chair, car, flower, and so on. It is also logically distinct from its identification (C) as a particular table, particular chair, etc. But there are obvious relations between (A), (B), and (C). The identification of a table as a particular table (C) logically presupposes its identification as an *individual table* (B), and its identification as an individual *object* (A). One can of course identify a table as an individual object without identifying it as a table, or chair, etc. (one may not know that it is a table, etc.); and *a fortiori*, without identifying it as a particular table or chair (as so-and-so's table or chair; and so on).

2. *The Identification of Individual Objects*

With these preliminary remarks we return to our working question of how in fact we identify an object as an *individual object*. The spatio-temporal "dis-

position" of an object—the place in which it is at a given moment or stretch of time—appears at first sight to supply us with part of the answer. For it is impossible for two objects simultaneously to occupy exactly the same place. This, we saw, is itself entailed by the ordinary meaning of "an object"—and we may now add, of "a place." Further (what we call) an object cannot simultaneously occupy two different places—again by the meaning of these expressions. Indeed, it appears that no kind of individual existing thing can be simultaneously in two different places.

We shall now briefly consider the ordinary uses of "(in) a place" and "(in) the same place," and then attempt to show that an object's occupancy of a certain place at a given time does not help us to distinguish it from other objects.

The phrase "a place" has a whole range of applications in ordinary discourse. A place may be very small or very large, the place occupied at a given moment by a very small or a very large object, or the place at which there is a small or a large surface, light (luminosity), etc. The coin before me is in a certain place on the desk, in the room (or relative to these things); and though a furnished apartment or building may contain many items of furniture, each of which is an individual object, we can speak of the apartment or the building, furniture and all, as in a certain place in town. However, although there is a use of "in a place" in which two or more objects—the furniture in an apartment, coins in a wallet, handbag, or cupboard—are ordinarily said to be in a certain place, we do not say that they are in *exactly* the same place (in the apartment, wallet, etc.). Things that are all in a certain place are so in a general sense of "in a place." In a more "precise" use of the expression, they are more or less, or roughly, in the same place. Similarly two or more objects do not and cannot occupy exactly the same place at the same time.

Are the different *parts* of an object at or in the same place? Do we ordinarily say that they are at or in the same place? I am inclined to think that, with regard to certain though not all kinds of objects (bodies, structures, etc.) we do *not* ordinarily apply to their parts the notions of "same place" and "different places." We speak only of the place of the object as a whole. This seems to be generally true of small or relatively small objects with continuous parts. On the other hand, the situation seems to be different with regard to (1) certain large or relatively large objects. We speak of the rooms, balconies, pillars, walls, ceilings, floors, etc., in a flat or house,

and the different floors of a building, as being in different places in the flat, house, etc., respectively. (We also say the rooms and balconies occupy different parts of the flat, etc.) Similarly (2) with regard to objects that can exist either separately or form part of larger wholes (some of the things we mentioned under (1) are examples of this). For example, each of the connected wagons of a train occupies a certain place different from the place occupied by the other wagons of the train. (But the train as a whole is also said to occupy a certain place. Furthermore, each of the wagons is in the same general place as the train as a whole.) Another example, but different from the preceding example in certain interesting ways, is a table or a chair. A chair as a whole is, at a given moment, in a certain place, legs and all. Now suppose we detach one leg, but keep it where it was before it was detached. It can now be regarded as a separate object, and can (therefore) be said to occupy a place of its own, slightly different from the place occupied by the (rest of the) chair. If, finally, we imagine the detached leg glued back to the chair, at the point at which it originally was glued, we would stop talking of it as occupying a place of its own, and speak only of the chair as a whole—legs and all—as occupying a certain place.

Again, (3) the different portions of such things as wires and pipes, which may be miles long, are I think ordinarily said to be in different places relative to one another or to other things. They are also said to occupy, at a given moment, different parts of space.

There are probably other, perhaps many other, types of cases of the present kind, all of which illustrate the complexity of the uses of "(in) a place" and "same place," as well as of "(spatio-temporal) part."

With this we turn to the putative role of spatial position in our identification of individual objects.

Point I. I said that two objects *O* and *P* cannot simultaneously occupy *exactly* the same place. Consequently if *O* is at or in a particular place p_1 at a certain time, *P* must be at or in a somewhat or totally different place p_2 at that time, using "place" in a very narrow sense; i.e., where "same place" means "exactly the same place." This does not mean

that "a place" is implicitly defined in ordinary discourse by the meaning or uses of "an object." The concept of a place is logically independent of the concept of an object; and a place can be identified by other means than the presence of an object or a number of objects at or in that place. For example, it can be identified by a light, an event, or even a sound.³⁵ However, if we simultaneously perceive what we correctly regard as *two objects A and B*, we can at the same time distinguish at least two different places: the different places *A* and *B* respectively occupy. Some places can therefore be distinguished (or identified) relative to the objects that occupy them. Also, the spaces between objects, whenever such spaces exist, exist and are identifiable relative to the objects between which they exist.

If the notion of a place in general can be defined without reference to objects, it would seem, *a fortiori*, that the actual place a particular object happens to occupy at a given moment can be identified independently of that or any other object. But obviously we must first be able to distinguish individual objects before we can identify a particular region of space as a place occupied by an object rather than as two places, occupied by two objects, etc.

Nevertheless, space and time serve negatively as a "principle of individuation." As we saw earlier, if what we shall call '*A*' does not coexist with what we shall call '*B*' (I use '*A*' and '*B*' merely as symbols, for purposes of reference) it necessarily follows that '*A*' and '*B*' cannot refer to the same individual thing. The same would be true if what we refer to by '*A*' and what we refer to by '*B*' are clearly separated by an (another) object or set of objects, or by a space, in the ordinary meaning of "a space." But the idea of *A*'s and *B*'s being separated by another object or other objects obviously presupposes the possibility of correctly identifying at least one object; and so spatial separation cannot be a logically ultimate criterion of numerical diversity. Similarly, *mutatis mutandis*, with the correct identification of a space.

Now we speak of moving an object *O*, or of *O*'s motion, from one place to another; and this frequently means *O*'s motion relative to other objects. For example, I may move a table in this room to another place in it, i.e., relative to the room as a whole. The concept of a place (as well as the concept

³⁵ Thus J. O. Urmson is right in his criticism of Strawson's tying of the notion of "same place" to "same material object" in *Individuals*. However, Urmson himself goes a little too far in his view that "It is no doubt true that places are recognized by the material objects which occupy them . . ."; though he immediately adds "but it does not seem to me to be a conceptual truth 'that places are defined by the relations of material bodies,' as Strawson claims on p. 58, unless of course we refuse to count any other way of determining place as determining place in the same sense of that word." (Review of *Individuals*, *Mind*, vol. 70, [1961], p. 260.)

of motion) necessarily involves the idea of a spatial frame of reference. But as I said before in other words, the frame of reference (a) need not be an object, and (b) need not always be the same object—when it is an object to begin with. For example, although in my individual experience I quite possibly utilize my own body as one of the original spatial frames of reference for distinguishing places, once I reach the abstract idea of a place I can think of the position of my own body relative to something else. Indeed, in the very act of locating an object *O* relative to my body at time *t*, I will have located my body relative to *O* at that time. (Compare and contrast motion.) Note that the abstract idea of absolute position in space is, psychologically and epistemologically, a later development. It is not itself a perceptual phenomenon.

I said that in our individual experience our body is quite possibly one of the original spatial frames of reference. More generally, we probably first “define” *particular* places, at a given moment, by reference to the various objects we perceive, relative to one another and/or to our body: we assign places to the objects we perceive at a given moment, relative to one another and/or to our body at that moment. More correctly, we simultaneously assign a position to our body and positions to the different objects in our field of vision or touch at a given moment, relative to one another. I might add that part of our body, but usually only a part of it, frequently forms part of our visual field. But at any given moment we are roughly aware where our body as a whole is relative to the contents of our visual field, and vice versa. And when we move our heads, or an object in our visual field moves, the contents of this field change partly or wholly; but that part of our body which is outside the visual field at that moment remains constant, as a spatial frame of reference.

Point II. We cannot validly say that certain qualities *a, b, c, d*, etc., which are actually qualities of an object *O*, are qualities of one object simply because they occupy the *same place*, while qualities *g, h, i, j*, etc., which are actually qualities of another object *O*, are *not* qualities of *O* but of another object, simply because they do not occupy that place. This is false apart from the semantic fact, mentioned earlier, that in ordinary language it makes no sense to speak of any quality *a* of an object *O* as being in or

at the same place as any of the other qualities, *b, c, d*, etc., of *O*. For though we can *see colored expanses* having certain shapes and sizes, and we sometimes *experience* a certain hardness, softness, smoothness, or roughness when our hands explore the same region of space in which (as we say) we have our visual perceptions, the perception of these qualities in the same region of space does not warrant our saying that we have located (a) an object as opposed to, say, a surface, and (b) a single object as opposed to two or more contiguous objects. On the other hand, the fact that we may perceive what may appear to be two sets of qualities in *two* contiguous regions of space does not necessarily mean that we have located two objects as opposed to one object.

Point III. The impossibility of two objects’ simultaneously occupying exactly the same place is logically related to the impossibility of an object’s—indeed, of any other kind of existing thing’s—simultaneously possessing “incompatible” qualities, such as being green and red, round and square, all over. If two objects could simultaneously occupy exactly the same place, then the same surfaces could be simultaneously green (or red, etc.) and not green (not red, etc.) all over; and the same volume of space could be occupied by “something” whose spatial parts are, at the same time, both smooth and rough, hard and soft. And so on. But if we see something we shall label *S* possessing incompatible qualities, we do not, on that score alone, necessarily assert that it consists of two or more distinct objects, each possessing one of the incompatible qualities. For *S* may be what we rightly call a single object, one *part* of which is, say, green and soft and smooth, and another part red and hard and rough. The question we are faced with then is: What features lead us to decide between these two alternatives, when we observe incompatible qualities in the “same general region of space”? The answer would provide us with the criteria-features of individuality with regard to objects; and any purported “principle of individuation” which fails to give a satisfactory answer to this question cannot serve as such a principle. Thus Leibniz’ celebrated principle of the identity of indiscernibles fails as a “principle of individuation”³⁶ with regard to actual objects (i.e., in its empirical form) because, among other things, it fails in this very respect. For supposing no two individuals (e.g., two objects) are qualitatively identical, the

³⁶ Which logically follows from the principle as an alleged true empirical proposition asserting the nonexistence of any numerically distinct individuals that are qualitatively identical. That is, it would seem that if this proposition is true, qualitative difference logically and “ontologically” serves to distinguish numerically distinct objects; and so appeal to such difference can in principle (it would seem) enable us to distinguish distinct objects.

qualitative differences between them are not sufficient numerically to distinguish them: they do not constitute a sufficient ground for distinguishing two objects *A* and *S* as opposed to two qualitatively different parts *a* and *b* of the same object *O*. *A fortiori*, Leibniz' principle does not provide a criterion for actually distinguishing two or more objects.³⁷ As we shall see, other factors must be considered to decide between these two possibilities in any given case.

We can conceive of a possible world in which the principle, in a certain form, would be true. In such a world qualitative difference would be both a necessary and a sufficient condition of numerical diversity: that world would contain no two qualitatively identical individuals. It would be interesting to inquire whether the world in question would be necessarily the kind of world Leibniz himself sketches, and which he (wrongly) identifies with the actual world. But this lies beyond the scope of the present discussion.

Let us now consider a chessboard with white and black squares. When we look at it, we sensibly perceive the different squares as distinct extended parts of the one continuous surface. Each square is clearly distinguished from the others because it has a different color (black or white) from the color of the squares immediately around it. The extendedness of each square and its being colored are closely related. For whatever is colored is extended; moreover, to perceive each of the areas as colored is to perceive it as extended. Something may, of course, be extended but be colorless. But if the "patches" are literally colorless, we would fail to perceive them—and so fail to perceive them as extended—precisely because we fail to perceive a color or colors. That is, we perceive nothing. An example is a pane of clear glass that we see through but do not see.

On the other hand, the squares on the chessboard could be perceived as *distinct* areas even if they all had exactly the same color, provided that there were observable *boundaries* between them; i.e., provided that the chessboard was divided by a series of horizontal lines intersecting a series of vertical lines. With regard to actual chessboards the boundaries between the different squares are formed by the meeting of areas which have different colors.

The upshot is that the spatial boundaries of any

part of the surface of the chessboard define it as a distinguishable *part* of the chessboard-surface as a whole. But the chessboard as a whole has its own *three-dimensional* spatial boundaries—in a somewhat different sense of "boundary"—distinct from the boundaries of other, including adjacent, objects. Our perception of these boundaries leads us to say that the chessboard is an object numerically distinct from other objects.³⁸

But what defines these boundaries? As I said, visual sense-perception of the geometrically self-enclosed outlines of the object, or certain tactual perceptions. I can run my hand along one side of the board and round the edge along the other side, and so on, until I return to my starting-point. But there is a further criterion of individuality: the fact that the chessboard as a whole can be *separately moved, manipulated*. It can be lifted, taken into another room, etc., independently of the other objects in the room: the other objects remaining where they were.

The parts of an object move together because of their connectedness to or continuity with one another. A table has a continuous surface, and its legs are connected to the top of the table; e.g., they are glued to it. So when I move the table, the whole table—legs and top—move. If, as I move it, one of its legs falls off, I say that it has become detached from the table. It is now a distinct object, spatially neither connected to nor continuous with the rest of the table.

Now suppose that some or all of the objects in my visual field, except my body, are whirling about with considerable speed. In that case I can only continue to identify, or can *reidentify*, these objects by one or more of their *qualitative differences*, such as their sizes, shapes, or colors; the sound which one may be emitting to the exclusion of the others, and so on. But as stated before, we cannot identify, and so cannot reidentify, each of these things at a given moment, by its being perceived in "a different position" relative to our body, including our line of vision.

The fact that an object can move or be moved relative to the rest of our visual or tactual fields, and/or relative to our body, seems to play an important role in our actually coming by the notion of (a) place, and, to that extent, the notion of a particular

³⁷ Qualitative difference is also not a *logically* necessary condition of numerical distinctness—contrary to Leibniz' view—with regard to the *ordinary* uses of "an object." See my "Relations" and "Individuals and the Identity of Indiscernibles," *Proceedings of the International Leibniz Congress* (Hanover, 1966.) See also my *The Coherence Theory Of Truth* (Beirut, 1961), pp. 94 ff.

³⁸ Many objects have what we might call a "closed" three-dimensional outline or closed boundaries; they are spatially completely separate from the other objects or things in their environment. But this is not universally true. It is also not true of such things as buildings, hills, mountains, which are individual things, though not objects in the relevant ordinary sense of the word.

place. As we say, the object was in one place and is now in another. But we could have the notion of a place if nothing in our surroundings ever moved and we were rooted to our place all our lives; and even if we could not move our heads or eyeballs. We could still distinguish different spatial parts of our visual fields, such as left, center, and right; and we could distinguish things as in front of us (our body). But objects would not be perceived as at different distances from us if all the objects we perceive are out of our reach, and if the tactile and motor sensations of grasping objects, etc., are necessary for the awareness of distance.

Finally, "independent mobility" is not, by itself, always *sufficient* to distinguish an *object* from other kinds of existing things, e.g., clouds. It is also not always sufficient to distinguish *an* object from a group of objects connected physically in one way or another. For example, each wagon in a train is a distinct thing ("object"); yet all the connected wagons move together when the train moves. (At the same time, a train is largely composed of wagons.) Again, it is not always necessary that an object be actually mobile in order that we may be able to distinguish it from other objects. A building or a mountain provides us with a simple example.

Point IV. In the fourth and last place in this discussion, and connected with what we said under Points I–III above, *linguistic convention* plays an important role in determining whether some occupant of space is (to be regarded as) part of an object but not also an object in its own right, an object, a collection of objects, or some kind of complex of objects. Thus what we call a wagon is, conventionally, both a thing in the ordinary meaning of "thing" (an object in our extended use of "object"), and part of what we call a train; which is (also) ordinarily regarded as a thing ("object"). Generally speaking, something which satisfies the conditions

we stated above—and these are clearly connected—but which may also form part of a larger thing, by being spatio-temporally connected to it in some manner, is conventionally regarded both as an individual thing and part of a larger or more complex thing. There are perhaps many other reasons why certain kinds of perceived things are referred to by distinct general names such as "table," "chair," "car," etc.; so that if something is correctly called a table or a chair, etc., it is implied that it is *an* object. The fact that something serves or can serve one or more functions or kinds of functions, that it has one or more uses or kinds of uses, is one of these (e.g., in the case of artifacts). What we refer to as different parts of such an object do not separately serve the particular function(s) or use(s): certainly none of them serves it (them) in the degree in which or the efficiency with which the object as a whole serves it (them). But each part helps the object to serve its particular function(s), etc. Again, artifacts are obvious examples. However, this is only generally true with regard to things that are spatio-temporally connected to one another.

Generally speaking, convenience in reference appears to be an important factor with regard to many objects; and still other factors are also probably involved. A study of the culture or subculture in which a particular language is the dominant means of verbal communication should reveal the various reasons for such things as the fact that certain kinds of objects but not others have specific common names in that language and the fact that some or all of the parts of certain kinds of objects have specific common names, while some or all the parts of other kinds of objects do not have any specific common names. Similar considerations hold with regard to the various kinds of collections, groupings, or complexes of objects.

III. KIERKEGAARD'S PSEUDONYMS: A NEW HIERARCHY

ALASTAIR MCKINNON

THIS paper describes a recent attempt to establish a hierarchy among eight of Kierkegaard's pseudonyms. It emphasizes the points of philosophical interest in this work and makes only minimal reference to our method and use of the computer. Those interested in these matters are invited to consult forthcoming reports dealing specifically with these aspects of this investigation.¹

Of course many Kierkegaard students will see this investigation as completely mistaken or, perhaps, intentionally ironical. They will point out that his works are addressed to the individual reader and that a merely scientific analysis must miss their real nature and purpose. Now I am not wholly unaware of the dialectics of the situation. I accept the distinction between the reader's and the don's point of view and my own preference is certainly for the former. But I am convinced that, so far from having to choose between the two, we can, with reasonable wit, use one to clarify and enrich the other. More specifically, I believe that we can use the computer to demonstrate that Kierkegaard's account of his pseudonymous authorship is substantially correct and must be taken quite seriously. In any event, this is what I propose to show. Irony there is but its real root is that Kierkegaard's own account has so often been forgotten.

I

While Kierkegaard acknowledged most of his writings he deliberately assigned many of his most important and celebrated works to one or other of his various pseudonyms. His account of these literary creations is extremely interesting. They are, he says, "poetic constructions" or distinct literary personal-

ities, each having his own characteristic life-view which he expresses with an ideal consistency.² Kierkegaard insisted upon the importance of his pseudonyms,³ noted that there were contradictions between their views,⁴ begged that anyone quoting from their works would cite the name of the responsible pseudonym,⁵ and even suggested that a current misinterpretation of one of his categories was due to the fact that these warnings were not taken seriously.⁶ Indeed, at one point he wrote "So in the pseudonymous works there is not a single word which is mine . . ."⁷ Such claims appear to suggest that the views of the pseudonyms bear no positive relation to the real Kierkegaard and that their writings may and perhaps ought to be disregarded. I would take a less radical line. Following a latter hint of Kierkegaard, I propose to order the more interesting pseudonyms in relation to their creator. This should enable us to take account of Kierkegaard's warnings and at the same time preserve the pseudonymous works as a source for the understanding of his thought.

The explanation of Kierkegaard's use of the pseudonyms lies, of course, in his aim and strategy. As he frequently insisted, his primary aim was not to articulate some doctrine or teaching but rather, through his works, to lead his reader to an authentic existence, to spiritual maturity or, in his most familiar formulation, to become a Christian. It was to present Christianity to his contemporaries who, he believed, lived in aesthetic or, at most, in aesthetic-ethical categories. This is the real source of indirect communication, the many pseudonyms and, in large measure, the doctrine of the stages. In fact, the real purpose of the pseudonymous authorship was to present a series of alternative life-views

¹ Cf., e.g., Alastair McKinnon, "A Method of 'Author' Identification," *Computer Studies in the Humanities and Verbal Behaviour* (forthcoming).

² Cf. "A First and Last Declaration," *Concluding Unscientific Postscript* (hereafter "Declaration").

³ Cf. "Declaration" and *The Point of View for My Work as an Author* (Oxford, 1950), *passim* (hereafter *Point of View*).

⁴ *Journals*, 1238 and "Declaration."

⁵ *Ibid.*

⁶ *Point of View*, p. 126.

⁷ "Declaration."

with a consistency impossible for any actual person and in such a way that, taken together, they presented a genuinely dialectical progression from one stage to the next. The pseudonymous authors are ideally suited for this role because they are perfect expressions of these alternative views; or, better, because they are simply their living or audible embodiment.⁸

The importance of such a hierarchy can be illustrated with reference to *Enten-Eller* which in this respect, as in so many others, is a kind of paradigm of the entire pseudonymous authorship. The first volume of this work champions the aesthetic mode of life in a way deliberately designed to express and convey this attitude: the second presents the ethical mode of life in a quite different but equally appropriate manner. But these views are here presented as mutually incompatible. Now as readers it is sufficient for us to feel this conflict and, as Kierkegaard intended, to make our own choice. This requires no knowledge or opinions about Kierkegaard; indeed, from this point of view, it is even important that we should neglect his personality. But as those seeking to understand this complicated and fascinating thinker we plainly cannot take this road. The pseudonyms canvass virtually every possibility and if we would reach the real Kierkegaard we must know which are closer to him and which more remote. We must, in fact, learn to plot the relations between Kierkegaard and his various pseudonyms.

While most commentators have tended to avoid this problem altogether, certain eminent Kierkegaard scholars have recently recognized its urgency and attempted to provide some kind of answer. Geismar,⁹ Hirsch,¹⁰ Billeskov Jansen,¹¹ and Thulstrup¹² have ordered some of the works often in broad categories and sometimes without much reference to the specific problem of the pseudonyms. I admire these efforts but am bound to say that these answers are greatly at variance with one another, that most are inconsistent with Kierkegaard's own clues and that some are, if I may say so, inherently implausible. Indeed, their real merit would seem to be that, coming from such authorities, they underscore both the urgency of the problem and the necessity of solving it upon some quite objective basis.

⁸ *Ibid.*

⁹ Ed. Geismar, *Søren Kierkegaard, Hans Livsudvikling og Forfattervirksomhed*, 11, *passim*.

¹⁰ E. Hirsch, *Kierkegaard-Studien*, 11, pp. 672 ff., 747 ff., *et passim*.

¹¹ F. J. Billeskov Jansen, *Studier i Søren Kierkegaard's litterære Kunst* (Copenhagen, 1954). Cf. also F. J. Billeskov Jansen, "Essai sur l'Art de Kierkegaard," *Orbis litterarum*, vol. 10 (1955), pp. 18-27.

¹² Neils Thulstrup, "Commentary," *Philosophical Fragments* (Princeton, 1962), pp. 146 ff.

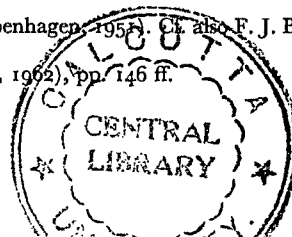
¹³ Cf. *Journals*, 641, 1000; *Repetition*, pp. 43 f., *et passim*.

Thus far we have described the pseudonyms as literary figures or creations—and this in keeping with Kierkegaard's own account or, perhaps, emphasis. It should be recognized, however, that he also thought of them in psychological terms. As he says, they are "possibilities" within his own nature. In fact, this conjunction sheds much light upon his life as an author. Real life usually allows us to become but one thing; it forces us to choose one direction or another. But the world of literature makes it possible for us to become all things; given the necessary talent, it allows us to develop each of our potentialities to its full perfection. It is this, of course, which lies behind Kierkegaard's conception of his authorship as a "therapeutic" exercise. In it he was exploring, trying out or, perhaps, trying on the possibilities he recognized within his own nature.¹³ This, of course, gives added point and another dimension to our study.

On Kierkegaard's own account, the pseudonymous works are significantly different from the acknowledged ones and, equally important, each is significantly different from the other. Again, on his own showing, some of his pseudonyms are "closer" to him than others. It may prove helpful to conceive our method as a procedure for evaluating in turn these two distinct claims.

II

Of course there are already a number of author identification methods which might have been adapted to solve our problem; those of Ellegård, Morton, and, especially, Mosteller and Wallace are but some of the possibilities. But there are clear considerations against the use of such methods in this particular case. While the pseudonyms appear clearly distinguishable from one another, it is perhaps too much to expect that each would have his own distinctive "thumb-print." Further, the distinction between the pseudonyms is not simply one of mere "style"; on the contrary, the pseudonymous works are best seen as a generally successful attempt to overcome all facile versions of the style/content dichotomy. As Kierkegaard suggests, the pseudonyms are literary personalities whose distinctive



life-style is so perfectly expressed in their work that what they say is finally indistinguishable from the way in which they say it.¹⁴ What is needed then is a method capable of taking advantage of this fact and, not being restricted to a few grammatical or filler words, able to provide a comprehensive picture of the different text selections; in short, a method possessing the formalism and rigor of statistics and offering as a by-product real insight into the nature of the works themselves. In fact, our procedure yields two or, in the case of the pseudonymous works, three separate vocabulary lists for each selection. Preliminary examination of these suggests that at least one may serve as a basis for a simple and useful abstracting procedure. As will appear, they have already provided an interesting confirmation of some of our statistical findings and to this extent support for our method.

Our method has two separate stages, each with its own role corresponding to the claims already mentioned. The first stage has two parallel parts; a comparison of the vocabulary densities of the pseudonymous (PS) and acknowledged (SK) sets and, secondly, a comparison of their internal coherence or homogeneity by means of the vocabulary connectivity method. The purpose of these comparisons is to show that, both individually and collectively, the PS selections are significantly different from the SK. The second stage of the procedure is a pair-vocabulary test the purpose of which is to establish a hierarchy of the pseudonyms in relation to the acknowledged Kierkegaard. These steps will be explained in greater detail later in this paper.

Our first task was the selection and preparation of text. We chose eight selections from the pseudonymous works (hereafter PS₁, PS₂, . . . , PS₈) and eight from the acknowledged ones (hereafter SK₁, SK₂, . . . , SK₈). As indicated in Table 1, all these selections were substantial and the sets total 496 and 483 pages respectively. Four of these selections represent an original work in its entirety; the remainder are carefully selected samples made necessary by the size or, in certain cases, the nature of the original.

One selection requires particular comment. Because there was no obviously right choice for the eighth acknowledged selection (i.e., SK₈), I constructed a deliberately synthetic creation consisting of three newspaper articles, a piece about an actress

and a sermon, the three parts ranging over most of Kierkegaard's relatively brief literary career. My intention was to produce a truly typical if fictitious selection which might, as it were, anticipate the acknowledged Kierkegaard. In fact, as our statistics suggest, this was based upon a misunderstanding and we have had to live with my mistake.

Each of these 16 selections was pre-edited by marking all nouns with a 1, all verbs with a 2, all adjectives with a 3, and all adverbs with a 4, a procedure permitting the subsequent recovery of all instances of each class of our content vocabulary. The entire text of each selection was key-punched in free format together with the markers for these four types of words. A master word list was produced and manually canonized (i.e., all variant forms of all words of these four types were reduced to their dictionary or root form). The computer was used to produce canonized word lists and rank-frequency distribution tables for each individual selection and for both sets. These tables were recorded for use in subsequent computation of the various calculated vocabulary connectivities. I can omit the remainder of these rather pedestrian details but would stress the following. Our method considers only these four types of content words and it is these which form the basis of our subsequent measures of text length and vocabulary size. Further, except where otherwise indicated, it deals with the vocabulary item as such as distinct from its various (word) forms or, indeed, the number of their occurrences. Thus, even the repeated occurrence of *høj*, *højere* or *højest* within a given selection is taken to mean only that this selection contains the vocabulary item *høj*. This said, it now remains to explain the steps of our method in greater detail.

The comparison of vocabulary densities is extremely simple and requires only the briefest comment. Through detailed empirical investigation we discovered that, for our sets at least, vocabulary size grew with text length according to a bi-logarithmic relationship. This means that the logarithmic ratio of vocabulary to text ($\log V / \log N$) represents an adequate measure of vocabulary density. We use this measure to compare the PS selections and set on the one hand with the SK selections and set on the other. This is the first step in the first stage of our method.

The degree of internal coherence or homogeneity within these two sets can also be measured by the

¹⁴ Cf. the following comment of Kierkegaard upon the pseudonymous works: "If you are capable of it, present the aesthetic with all its fascinating magic, enthrall if possible the other man, present it with the sort of passion which exactly suits him, merrily for the merry, in a minor key for the melancholy, wittily for the witty, &c." *Point of View*, p. 29.

vocabulary connectivity method. For this purpose we establish the vocabulary connectivity, both calculated and observed, for both sets taken separately. In earlier versions of this work the calculated values were produced from a formula employing the number-frequency distribution of the set and the text length of the individual selections. This was found to be inadequate and in the present version these values are derived instead from a naive model or formula incorporating both individual text lengths and the rank-frequency distribution for each selection. Hence this model, unlike the earlier one, takes account of the differing vocabulary densities. The application of this formula gives the number of vocabulary items which would belong exclusively to a given selection on the basis of this model. The corresponding observed values were produced by the computer by comparing the vocabulary of all selections within each set. As this is a test of approximation to a purely chance distribution, the differences between the calculated and observed values in each case indicates the degree of homogeneity within each of these two sets.

The next and final stage in our method is the application of the pair-vocabulary test. This involves three distinct operations. The first is the determination of the number of observed vocabulary items each PS selection has *exclusively* in common with SKn. The second is the calculation of the number of such items which might be expected on the basis of chance alone using our own naive model for the nine selections (i.e., PS1, PS2, . . . , PS8 and SKn) and hence correcting for the differing lengths and densities of each selection. The third is the finding of the ratio between the observed and calculated values. Stated in very general terms, the assumption underlying this test is that the pseudonym whose vocabulary approximates most closely to that of the acknowledged Kierkegaard is therefore at the top of the hierarchy.

III

The statistics concerning vocabulary density (or richness) are given in Table 2. The text lengths of the SK and the PS sets are of the same order of magnitude and, this being so, we assume that they constitute comparable samples. We also assume that the individual selections are sufficiently large to permit the use of the $\log V / \log N$ ratio as a means of

comparison. Now, as this table shows, the ratio for both the individual PS selections and for the PS set is in general much higher than that for the SK selections and set. Indeed, the average for the PS set is .845 while that for SK is .825 or, deleting our synthetic SK8, .819. In brief, the PS authors have a much richer vocabulary than SK. Now if, with Somers,¹⁵ Baudot,¹⁶ etc., we regard this ratio as giving an important formal characteristic of a literary work, we have here clear evidence of a significant difference between these two sets. For my own part, I regard it as strong evidence for the claim that the works of the PS authors are quite distinct from those of the acknowledged SK.

This difference can be expressed somewhat more simply. The PS set has a total text length of 84,111 occurrences (of content words) and a total vocabulary of 8,466 items while the SK set has 82,670 occurrences and 6,562 items. Hence the vocabulary range of the former is approximately 29 per cent greater than that of the latter. In short, Kierkegaard has a much greater vocabulary range writing in the guise of the pseudonyms than doing so in his own person.

This same general conclusion is suggested by the vocabulary connectivity method which is in fact a parallel or second step in the first stage of our method. This involves a comparison of the observed with the calculated values for both the PS and the SK sets taken separately. More precisely, it involves a comparison of the number of vocabulary items which are in fact peculiar to a given selection within a set with the number of such items which might be expected on the basis of pure chance. The results of this comparison are given in Table 3.

This table shows that the SK set is much more coherent or homogeneous than the PS. This is most evident from a comparison of their respective differences. The magnitude of the differences for the SK set is 346 while that for the PS set is 1,312. Further, of SK's total of 346, 216 or approximately two-thirds come from our synthetic SK8. If this difference were removed the total of SK differences would drop to a mere 150. Clearly, therefore, the SK selections are in general much more like one another than the PS.

It is perhaps worth pointing out that whereas the SK differences contain an approximately equal mixture of positive and negative values, all the PS differences, except that of PS8, are high and,

¹⁵ H. H. Somers, "Statistical Methods in Literary Analysis," University Press, 1966), pp. 128-140.

¹⁶ Jean A. Baudot in recent private correspondence.

The Computer and Literary Style, ed. by Jacob Leed (Kent State

particularly, positive. This means that seven of these eight selections have a substantially larger number of exclusively vocabulary items than are predicted by our chance model. In short, these selections have a markedly large and rich vocabulary. Of course, this is also implied by the ratios in Table 2.

It would perhaps be reassuring if we had other, independent evidence of the adequacy of our naive or intuitive model but unfortunately this is not the case. We can, however, point out that, even with the overall distortions resulting from the inclusion of SK8, it was nevertheless able to predict the vocabulary connectivity values for the remaining seven acknowledged selections with remarkable accuracy. But the crucial point is that this same formula was used in both cases with two very different results. Whereas it showed the SK set to be truly coherent (as one would expect), it also showed the PS set to be clearly incoherent. It seems, therefore, reasonable to conclude that the pseudonymous selections are, in the sense which Kierkegaard intends, from different hands. Certainly their vocabularies are so disparate that, in other circum-

stances, one would be tempted to regard each as the work of a different author.

This is perhaps the point at which to comment upon another possible use of the vocabulary connectivity part of our method. Originally we intended it both as a means of comparing the internal coherence of the two sets and as a basis upon which to discard any SK selection which proved atypical; this on the assumption that SKn should not contain any selection with a significantly large number of anomalies. However, several statisticians, perhaps assuming that we intended only the first use, have objected to the second. In spite of the evidence we have therefore decided, for the moment at least, to retain SK8 as part of SKn.

The final step in our method is the pair-vocabulary test. As already suggested, this consists of a comparison of the number of vocabulary items each PS selection has exclusively in common with SKn with the number of such items it might be expected to have on the basis of our formula. These values, together with their ratios, appear in Table 4. The resulting hierarchy is given below together with the final ratios for easy comparison.

Selection	Ratio	Pseudonym	Work
PS8	1.0087	Anti-Climacus	<i>Sickness Unto Death</i>
PS1	0.8854	A	<i>Either-Or</i> , vol. 1
PS7	0.8682	Johannes Climacus (2)	<i>Concluding Unscientific Postscript</i>
PS2	0.8426	B	<i>Either-Or</i> , vol. 2
PS6	0.8073	Johannes Climacus (1)	<i>Philosophical Fragments</i>
PS3	0.7546	Johannes de silentio	<i>Fear and Trembling</i>
PS5	0.7134	Vigilius Haufniensis	<i>Concept of Dread</i>
PS4	0.6047	Constantine Constantius	<i>Repetition</i>

IV

Of course, it would be possible to produce still further refinements in this method; indeed, I can think of a few myself and, given time and resources, might even be tempted to try some of them. Nevertheless I believe that the present version is substantially sound and that we can safely proceed to consider our results and their consequences.

Clearly the most obvious, certain, and important consequence of this work is that it provides an independent confirmation of Kierkegaard's account of his entire authorship and, particularly, of the nature and status of the pseudonymous works. Both Tables 2 and 3 show that all the acknowledged selections apart from SK8 are extremely homo-

geneous and, in terms of these measures, are rightly regarded as the work of a single author. (This is doubtlessly also true of the individual selections comprising SK8, the aberrant scores of which are almost certainly due to its synthetic nature.) On the other hand, all our evidence shows that the pseudonymous selections are quite unlike the acknowledged ones and, equally important, quite distinct from one another. Nor is it simply that, as Kierkegaard pointed out, each pseudonym has his own distinctive life-view; these results show that each also has a substantial distinctive vocabulary in and through which he expresses that view. Indeed, they show that each pseudonym is, as Kierkegaard said, a distinct literary personality. They show that Kierkegaard's warnings concerning his authorship

are entirely justified and that there can no longer be any excuse for not taking them seriously.

The importance of these warnings can be illustrated with an example. *Fear and Trembling* is a widely known and popular work defending what it calls "the teleological suspension of the ethical." Now despite Kierkegaard's warnings about the status of the pseudonyms, despite his own personal repudiation of this particular view, the majority of commentators continue to ascribe it to Kierkegaard and many even seem to make it central for their interpretation of his thought. But our results show plainly that this will not do. They show that the pseudonymous works constitute a class by themselves and, further, that within this class this particular work has no such privileged status. Indirectly at least they support Kierkegaard's claim that each of the pseudonymous works must be seen as a poetic statement set within the context of his overall strategy.

We can perhaps see something of the reality and importance of the distinction between the pseudonymous and the acknowledged works from the fact that, according to our lists, the words *Paradoks* and *Absurde*, together with their variants, occur many times in the pseudonymous works but never in the acknowledged ones.¹⁷ This is more remarkable in view of the fact that these occurrences are almost always triggered by some reference to Christianity, which references are in fact much more numerous in the acknowledged works. It follows from this that the characteristic conception of Christianity which informs the pseudonymous works is significantly different from that which informs the acknowledged ones. It also follows that there are certain important advantages in counting content rather than filler words.

We have seen that certain vocabulary items are assigned to the pseudonymous works as such but the matter is in fact yet more complicated. Many others appear to be clearly and specifically reserved for some particular pseudonym. Thus, for example, our relevant list gives *Opoffrelse* (sacrifice) as a pair-vocabulary item for PS8. But this means that this particular word does not appear even in *Fear and Trembling*, a fact so remarkable that one is virtually forced to postulate the operation of something like conscious choice and repression. However, this is but one example. Table 3 shows that each of the pseudonymous selections has a very large number of

vocabulary items peculiar to itself just as Table 4 shows that each has a surprisingly large number which it shares exclusively with SKn. I have not yet been able to do a detailed analysis of these lists but preliminary examination of the pair-vocabulary of two of the pseudonyms suggests that at least three-quarters of these items have no apparent tie with the subject under discussion. In short, the pseudonyms, like the rest of us, have a certain number of characteristic vocabulary items which continually reveal their identity. Indeed, one begins to suspect that, like most writers who employ pseudonyms in the more conventional sense, they apparently cannot resist the temptation to give themselves away with certain fairly obvious clues.

Lest the reader remain entirely sceptical about this approach I should perhaps report the following. At an earlier stage in this investigation, working with another formula, a different correction factor and a reduced SKn, we discovered that PS1 emerged clearly at the top of our hierarchy. This result so surprised us that we resolved to see whether it might be due to the possibly atypical "Diapsalmata," a series of brief and apparently unconnected fragments used as the first one-third of this selection. In fact, our manual count revealed that of the 208 pair-items, 69 came from the first one-third and 139 from the remainder. In short, in terms of our measure as applied under these conditions, even this selection was constant throughout.

Before considering our hierarchy in detail it is worth noting that it does not follow any of the obvious orderings of these works. It is not a function of similarity of subject matter to that of the acknowledged works; that would have required, for example, that PS3 and PS5 be at least near the top and that PS1 be at the bottom. Neither is it simply a function of vocabulary density or richness; PS1, with the richest vocabulary, is in second place while PS4, with the second richest, is at the bottom. Nor does it appear to be in any sense a function of the date of composition; PS8 is the latest and PS1 the earliest of all our pseudonymous selections and while it is possible to see the remainder as repeating this "zeroing in" effect, the mean date of the acknowledged works is too late to make this a plausible hypothesis. As these appear to be the only possibilities I would therefore argue that our hierarchy is precisely what we intended it to be, *viz.*, an ordering of the pseudonyms in terms of their proximity or

¹⁷ Since this was written M. Ferdinand Ouellet has pointed out that *Paradoks* occurs once in "The Instant" (Number 5) and a number of times in *On Authority and Revelation*. I am grateful for this observation but would still wish to maintain that these concepts have little if any place in the conception of Christianity presented in the main acknowledged works.

likeness to the real or acknowledged Kierkegaard.

I am fairly confident of the broad outlines of our hierarchy but would strongly caution against attaching too much importance to certain of its particular juxtapositions. Our method is a merely statistical one and differences between the ratios of our selections can be regarded as significant only if they are sufficiently large that they could not be expected upon the basis of mere chance. For the sake of clarity I would therefore suggest that difference be regarded as significant only if the first ratio is at least five per cent greater than the second. This means that not all our placings can be regarded as established beyond question. It should be pointed out, however, that this is not due to any deficiency of pair-vocabulary in the selections in question; the difficulty is simply that their ratios are not sufficiently different to justify the placing of one clearly above the other.

Though these limitations follow from statistical considerations they are also quite in keeping with the nature of our problem. In this connection there may be value in quoting from *Repetition* which, though it falls to the bottom of our hierarchy, plainly expresses an important and, I think, fundamentally Kierkegaardian insight:

"... but the individual has a multiplicity of shadows, all of which resemble him and for the moment have an equal claim to be accounted himself."¹⁸

"So does the possibility of the individual stray at random amongst its own possibilities, discovering now one and now another."¹⁹

Faced with these salutary reminders we can perhaps be content with a method which treats personality as a series of dynamic, fluid, and not altogether determined possibilities and which forces us sometimes to think in terms of relative priorities rather than definite placings. Indeed, it is possible to see these psychological truths as support for the use of a statistical method.

Expressed in terms of the pseudonyms, our hierarchy is as follows. Anti-Climacus is clearly in first place and Johannes *de silentio*, Vigilius Haufniensis, and Constantine Constantius clearly in sixth, seventh, and eighth place, respectively. On the basis of our own requirement that the difference must exceed five per cent we must regard A and Johannes Climacus (2) as tied for second place, Johannes Climacus (2) and B as tied for third, and B and

Johannes Climacus (1) as tied for fourth. By this same standard we can, however, say that A is clearly above B and that Johannes Climacus (2) is clearly above Johannes Climacus (1). Such priorities, though perhaps less satisfying, are at least as useful as absolute but perhaps mistaken placings.

Of all these findings, the most certain is that of the primacy of Anti-Climacus. Indeed, if I might disregard some of my own warnings for a moment, I would point out that this pseudonym has an even slightly larger pair-vocabulary than that predicted by our chance model. In fact, neither this nor his primacy is at all surprising. The theme and tenor of his work is sensibly close to that of the acknowledged Kierkegaard. The text lacks the indirectness of most of the other pseudonymous works and, indeed, the name *Anti-Climacus* was appended only as an afterthought. Further, Kierkegaard himself later pointed out that Anti-Climacus was, as his name suggests, one of the "antitheses" within his own nature.²⁰ We can therefore be quite certain that Anti-Climacus belongs at the top of the hierarchy and that he expresses an element or possibility which was very important within Kierkegaard's nature. Certainly, his primacy gives added substance to Kierkegaard's repeated claim that he had been "from first to last a religious author."

The next result indicated by our evidence is that A stands above all the remaining pseudonyms save, perhaps, Johannes Climacus (2). This is much more surprising except perhaps to a psychiatrist. In fact, this placing of A runs counter to all the obvious evidence. This pseudonym is the champion of the aesthetic way of life, a way which in Kierkegaard's life and thought is often in virtual opposition to the religious view of his acknowledged works. But the evidence is plain and incontestable. Despite the clear differences in their subject matters, A's pair-vocabulary ratio is significantly higher than that of any of the remaining pseudonyms save perhaps Johannes Climacus (2). But even if there is a tie between these two it is still clear that A is much closer to the real Kierkegaard than either he or the commentators have generally assumed. Put another way, what Kierkegaard called the aesthetic apparently is a very important element in his thought and personality; certainly it is very prominent among its "possibilities" or incipient forms. This may well be why Kierkegaard's account of Christian belief within the pseudonymous works

¹⁸ *Repetition*, p. 43.

¹⁹ *Ibid.*, p. 44. Cf. also *The Present Age* (Oxford, 1949), p. 51.

²⁰ See *Journals*, 936 and, especially, 1000.

is so notably shrill. Perhaps, as he seems to have suspected, he was also naturally an aesthetic man after all.

The ascendancy of A over B, and hence of the aesthetic over the ethical, is itself worthy of particular comment. Though at variance with traditional accounts of the stages, this ordering is not at all implausible. The pseudonymous spokesman for the ethical is in many ways less attractive and compelling than the champion of the aesthetic and it is often clear from the tone of the former that Kierkegaard's heart is not in him. There is also the fact that he personally later repudiated the ethical as pictured within the pseudonymous works. Indeed, this conception of the ethical appears to reflect his understanding of his own personal situation as seen through Kantian and, in another sense, Hegelian eyes and no doubt is partly determined by the demands of the authorship as he then understood it. It is clear, however, that his own conception of the ethical is specifically Christian and hence distinctly different; witness, for example, the views set forth in his own *Purity of Heart* and, especially, *Works of Love*. It is therefore not surprising that the ethical as conceived within the pseudonymous authorship should in fact be further removed from Kierkegaard than the aesthetic. Quite obviously, the latter was at least his own.

Neither, on reflection, is it surprising that Johannes Climacus (2) is significantly closer to Kierkegaard than Johannes Climacus (1). Of course, both pseudonyms bear the same name and both propound essentially the same thesis; indeed, their works are explicitly presented as original and sequel or postscript. Nevertheless there are clear and obvious differences. The argument of the *Fragments* is formal and at times almost algebraic while its mood is often somber and, on occasion, not unlike that of *The Concept of Dread*. The *Postscript* on the other hand is gay and even hilarious and has an inner momentum much like that of certain parts of the first volume of *Either-Or*. Further, the author of the *Postscript* is permitted to survey and comment upon the whole of the preceding pseudonymous literature. Plainly, then, he is a separate and privileged character and we should not be surprised that our method discriminates between him and his predecessor.

All of the remaining differences are significant and the subsequent placings hence well founded. These are, in order, Johannes *de silentio*, Vigilius Haufniensis, and Constantine Constantius. These three can be seen as standing at successive removes

from the real Kierkegaard. However, instead of dwelling upon these three I shall attempt to place the whole hierarchy in a broader perspective.

In very large measure the pseudonymous authorship is informed by the doctrine of the stages, a doctrine which Kierkegaard repeatedly modified in two quite separate respects. Very briefly, the first version postulates three stages, the aesthetic, the ethical, and the religious while a later increases the number to four by distinguishing between Religiousness A (the religion of immanence) and Religiousness B (the religion of transcendence or Christianity). In both cases, however, the various stages are seen as exclusive and, from one point of view, successive. In fact, the reader is being invited to move from one stage to the next. Later, however, Kierkegaard came to a quite different version which, for convenience, I shall call the "dethronement" view. According to it, there are but two stages and hence, as he says, but one either-or: *either* the aesthetic as such *or* Religiousness B, including within it, as dethroned stages, Religiousness A, the ethical and the aesthetic. But this is really what our own hierarchy suggests. It gives pride of place of the spokesman for Religiousness B followed, perhaps, by the champion of the aesthetic. These are followed, in turn, by the spokesmen of the other stages and, next, by Johannes *de silentio* who, so far from being spokesman, is the arbiter of a debate between the ethical and the religious. Finally, and at a considerable difference, come Vigilius Haufniensis and Constantine Constantius, authors sharing certain of Kierkegaard's concerns but having no explicit connection with any of the stages. Hence our hierarchy is confirmed by and may be seen as a reflection of Kierkegaard's final version of his doctrine of the stages and, equally, of the centrality of its focus for his own thought. The "stage" pseudonyms stand at the top of the hierarchy precisely because, like the acknowledged works, they are concerned above all else with what it means to exist.

No doubt results such as those we have produced can be used to formulate a more precise picture of Kierkegaard both as author and as person. No doubt such results can be used to decide whether Johannes Climacus really is one of the antitheses within Kierkegaard's character as he believed or whether, as our evidence seems to suggest, this honor perhaps belongs to the spokesman for the aesthetic. However, before making any final pronouncements on such important matters, I should want to check through our method again very thoroughly; in particular, I should want to be quite sure that the relative

primacy of A is not in fact partly due to the unfortunate inclusion of the clearly atypical SK8. Until these matters are settled, I should not want to say any more than has already been said.

For the moment at least the results of all this labor are truly miniscule. After many months and much expense we have been able to show that the pseudonymous works differ both from the acknowledged ones and from one another. We have also

been able to show that Kierkegaard's notions regarding priorities among the pseudonyms are in general substantially correct. Of course, it is ludicrous that anyone should have to use a computer to show that Kierkegaard understood the structure of his authorship but, as he often remarked, where illusions are deep-seated and wide-spread, a certain kind of deceit is sometimes necessary.²¹

Received April 23, 1968

McGill University

²¹ This investigation was made possible through grants from the Canada Council and McGill University, to both of which I express my thanks.

APPENDIX

Table 1 Selections from Acknowledged Works

Selection	Sample	No. of pages	Title	Date
SK ₁	5 of 18 Discourses	72	Atten opbyggelige Taler (Edifying Discourses)	1843
SK ₂	Complete	78	Tre Taler ved Tænkte Leiligheder (Three Discourses on Imagined Occasions)	1845
SK ₃	Complete	51	Hvad vi lære af Lilierne paa Marken og af Himmelens Fugle (What we learn from the Lilies . . . of the Air)	1847
SK ₄	Every second page	55	Lidelsernes Evangelium (The Gospel of Suffering)	1847
SK ₅	4 of 5 Sections	67	Kjerlighedens Gjerninger (Works of Love)	1847
SK ₆	1 of 4 Sections	75	Christelige Taler (Christian Discourses)	1848
SK ₇	Complete	34	Lilien paa Marken og Fuglen under Himlen (Lilies of the Field and the Birds of the Air)	1849
SK ₈	{ 3 articles	15	Bladartikler (Newspaper articles) Krisen og en Krise i en Skuespillerindes Liv (Crisis in the Life of an Actress) Guds Uforanderlighed (God's Unchangeableness)	1843—48 1848
	{ Complete	21		
	{ Complete	15		

Selections from Pseudonymous Works

Selection	Sample	No. of pages	Title	Date
PS ₁	Diapsalmata and Forførerens Dagbog (Every fourth page)	66	Enten-Eller. Første halvbind (Either-Or, vol. 1)	1843
PS ₂	Ligevægten . . . Udarbejdelse	52	Enten-Eller. Andet halvbind (Either-Or, vol. 2)	1843
PS ₃	Every second page	51	Frygt og Bæven (Fear and Trembling)	1843
PS ₄	C.C.'s sections complete	58	Gjentagelsen (Repetition)	1843
PS ₅	Every third page	43	Begrebet Angest (Concept of Dread)	1844
PS ₆	Complete	84	Philosophiske Smuler (Philosophical Fragments)	1844
PS ₇	Every sixth page	86	Afsluttende uvidenskabelig Efterskrift (Unscientific Postscript)	1846
PS ₈	Every second page	56	Sygdommen til Døden (Sickness Unto Death)	1849

Table 2 Text-Vocabulary Relationship

	Text Length N	Vocabulary V	$\log V / \log N$
SK ₁	11181	2176	0.825
SK ₂	14408	2807	0.829
SK ₃	8588	1732	0.823
SK ₄	10114	1940	0.823
SK ₅	11285	2074	0.818
SK ₆	12845	2180	0.810
SK ₇	6548	1262	0.813
SK ₈	7701	2144	0.860
PS ₁	10519	3132	0.869
PS ₂	9380	2258	0.845
PS ₃	8639	2229	0.851
PS ₄	9771	2868	0.865
PS ₅	7759	2057	0.851
PS ₆	13549	2509	0.824
PS ₇	15432	3032	0.831
PS ₈	9062	1854	0.827

Table 3 Vocabulary Connectivity Values

	Vocabulary exclusive to given selection		
	Observed	Calculated	Difference
SK ₁	447	412	+35
SK ₂	664	649	+15
SK ₃	268	281	-13
SK ₄	302	337	-35
SK ₅	352	375	-23
SK ₆	407	404	+3
SK ₇	170	164	+6
SK ₈	644	428	+216
PS ₁	944	583	+361
PS ₂	406	315	+91
PS ₃	383	311	+72
PS ₄	669	495	+174
PS ₅	401	274	+127
PS ₆	522	362	+160
PS ₇	765	516	+249
PS ₈	339	218	-121

Table 4

	Exclusive Vocabulary		
	Observed	Calculated	Ratio
PS ₁	269	304	0.8854
PS ₂	139	165	0.8426
PS ₃	123	163	0.7546
PS ₄	156	258	0.6047
PS ₅	102	143	0.7134
PS ₆	155	192	0.8073
PS ₇	237	273	0.8682
PS ₈	116	115	1.0087

IV. ARISTOTLE AND OXFORD PHILOSOPHY

RICHARD R. K. SORABJI

AMONG Aristotle's works, the *Nicomachean Ethics* has long been a favorite at Oxford. It is the text most commonly read by Greats men. And it has influenced work by Austin, Ryle, and Anscombe. One explanation for its traditional popularity was offered by Hastings Rashdall in *The Universities of Europe in the Middle Ages*.¹ Speaking of Robert Grossetête, he wrote, "The first translation of the *Ethics* direct from the Greek was made under his directions, though he can hardly be considered the actual translator: and it is tempting to trace back the peculiar and undying affection of Oxford for the *Nicomachean Ethics* to the influence of her first recorded Chancellor." But what explains the book's present popularity? There are many answers. One important factor, of course, is that it is a very good work of philosophy. But this is true of other writings of Aristotle too. The answer that interests me here is that in certain ways the *Nicomachean Ethics* is one of the texts most in harmony with recent Oxford philosophy. I do not have in mind the use that it makes of the data or ordinary language, which has been brought out by G. E. L. Owen (*τιθέσθαι τὰ φαινόμενα*).² For, as Owen shows, this feature is also

present in Aristotle's *Physics* and elsewhere. Rather I have in mind that certain major contrasts between Aristotle and recent Oxford philosophy that are noticeable in other works are absent or unobtrusive in the *Ethics*. Let us enumerate these contrasts.

I

(i) First of all, there is Aristotle's tendency to include in his definitions obscure and little known scientific information. Sometimes he puts a premium on the definitions which include this information, and allows it to be omitted only for special purposes (such as dialectic or studies in rhetoric). Examples are his definitions of thunder (*Posterior Analytics* 93b12-14), lunar eclipse (*Posterior Analytics* 90a15; 93a30-33; 93b6-7; *Metaphysics* 1044b12-15), musical consonance (*Posterior Analytics* 90a18-19), color (*On the Soul* 419a9-11; *De Sensu* 439b11-12), squaring the rectangle (*On the Soul* 413a13-20), the rightness of the angle in a semicircle (*Posterior Analytics* 94a34-35); being angry and mental occurrences in general (*On the Soul* 403a25-b9. Cf. the definition of a saw, *Physics* 200b4-8).³ One

¹ Oxford, 1895, vol. II, p. 521.

² *Aristote et les Problèmes de Méthode*, 2nd Symposium Aristotelicum, ed. by S. Mansion (Louvain and Paris, 1961).

³ See also the definitions of light (418b16-17); of hot and cold (329b26-29); of flavor (441b19-21); of odor (443a7). But it is not so clear that he would put a premium on these definitions.

Thunder and *lunar eclipse* are best defined by reference to their causes, which in Aristotle's time were still a matter for debate. One may even mention the cause of the cause. *Consonance* is not to be defined as the ability of two notes to blend into a single whole. Rather, the preferred definition will state the supposed cause of this ability, namely the simple numerical ratios which had been discovered to hold between the string lengths of any pair of consonant notes. The definition that gives the essence of *color* says that it stimulates the light between the eye and the object seen. Another definition calls it the boundary in a rigid body of the transparency, which in Aristotle's view exists in any body containing the elements air and water. In *squaring a rectangle*, one shows how to construct with ruler and compasses a square equal in area to a given rectangle, and finishes with a proof that it is the required figure that has been constructed. The preferred definition of squaring the rectangle will not only say that the constructed figure is equal in area to the rectangle, but will mention one of the methods of proving that it is, viz., showing that the figure's sides are of a mean proportional length. The favored definition of the *rightness of the angle in a semicircle* also picks out a possible proof, a proof which explains why the angle is a right angle, by showing it is half of two right angles in a certain figure. In defining *anger*, it is desirable to mention the boiling of the blood around the heart, which was supposedly involved. And corresponding physiological processes will be given in the definitions of other mental occurrences. One definition of a *saw* says its teeth are of iron. One definition of *light* calls it the presence of fire or of some such thing in what is transparent. *The hot* can be defined as what brings together things of the same kind, and *the cold* as what brings together both things that are like each other and things that are unlike. *Flavor* is an effect which is produced in what is fluid by a certain kind of dry stuff, and which can activate the sense of taste. *Odor* is the character of flavored dry stuff in fluid.

An important feature of many of these definitions, and one to which I shall recur, is that they include information which explains either the existence of the thing being defined (for example, it is explained why thunder occurs), or the existence of further features of the thing being defined (as the boiling of blood might explain some of the symptoms of anger). The first kind of case will be discussed in (iii)-(vi) of Section II, the second in (ii) of Section I and in (i) of Section II.

should notice also his view that a statement of what a word signifies is in some cases insufficiently informative as a definition (*Posterior Analytics* 93b29-94a14), and his view that rather than mentioning things known to us, one should mention things which are more properly objects of knowledge, if faced with a choice between these two courses (*Topics* VI.4). This last goes against a preference, frequently expressed in recent times, for mentioning what is known to users of the word being defined.⁴

(ii) One should notice not only the information that Aristotle includes in his definitions, but also the role that he wants them to play in a scientific system.⁵ He wants them to serve as the premisses of syllogisms. From the presence of the defining characteristics in the thing defined he wishes to deduce and explain the presence of other properties. For example, he mentions in the definition of anger his theory that it involves the boiling of the blood around the heart. Then from the definition he can infer and explain the heat and redness which is characteristic of angry people. In a similar way, one can use a definition to infer and explain further properties of the definiendum, if one defines thunder or lunar eclipse by reference to their causes, consonance in music by reference to the lengths of vibrating strings, or (better) by reference to the frequency of vibration, lemons by reference to a certain D.N.A. structure, iron by reference to a certain atomic number, heat by reference to atomic motion, and diseases by reference to particular kinds of viruses. These examples should help one to see that it is not an outlandish project to try to frame definitions that can be used for explaining and syllogistically deducing further properties of the thing defined. The sort of scientific project Aristotle had in mind was in some ways not unlike the very successful systems set up by Euclid a generation later, in a variety of subjects, geometry, arithmetic, acoustics, optics, and astronomy. If it is to be compared with anything in modern philosophy, it is perhaps most like the work of Carnap, a philosopher who has not had a large following in Oxford. One of Carnap's aims, as he explains in ch. I of his book, *Logical Foundations of Probability*, is to give definitions that will introduce in place of pre-scientific concepts new ones which are exact and which will facilitate scientific generalizations. For example, suppose the creatures which fall under the ordinary concept of

fish have fewer properties in common than the creatures that live in water, are cold-blooded vertebrates, and have gills throughout their life. Then a definition will be given, to introduce the new concept, defined by these latter properties, which will allow more generalizations. This aim of giving definitions that will facilitate scientific generalizations may be compared with Aristotle's desire to frame definitions that will facilitate deductions and scientific explanations. Carnap's readiness to improve on ordinary concepts also has its parallel in Aristotle, judging from such passages as *Posterior Analytics* 98a13-19, and *Parts of Animals* 669b9-13, and judging from his redefinition in *On Generation and Corruption* of the ordinary-language concepts of hot, cold, fluid, dry, discussed in *Parts of Animals*.

(iii) The amount of scientific speculation in Aristotle's writings is an obvious respect in which they differ from much recent philosophy. It is worth noting too the kind of importance that he attaches to his scientific work. He draws a contrast between theoretical *episteme* and dialectic. Theoretical *episteme* culminates in the kind of deductive system we have just described. It is divided into three branches, the mathematical sciences, the natural sciences, and metaphysics (*Metaphysics* VI.1). It includes both what we should regard as science and what we should regard as philosophy. Dialectic, on the other hand, would be classed by us as almost entirely philosophical, judging from the examples of dialectical reasoning in Aristotle. It is spoken of as inferior to theoretical *episteme*. It has a subservient part to play within theoretical *episteme*, where it helps one to reach the first principles of the individual sciences, or to reach the common principles which guide reasoning in all the sciences alike (*Topics* I.2, 101a37-b4; *Sophistical Refutations* 11, though it seems later to be stripped of the latter role, *Metaphysics* IV.3, 1005a19-b1; XI.4, 1061b17-33). And it has a bigger role in the disciplines which are classed as not theoretical, but practical (e.g., ethics, political science), or productive (e.g., poetics). But these disciplines are less lofty (*Metaphysics* 982a1; 1026a22-23; 1064b1-6; *Nicomachean Ethics* 1141a18-22; a33-b3; 1143b34; 1145a6-11). Theoretical *episteme* is the worthiest pursuit (*Nicomachean Ethics* 1144a3-5; X.7 and 8; *Metaphysics* 1072b24). The net result of this evaluation is that certain philosophical studies have been given a

⁴ See J. J. C. Smart, "Sensations and Brain Processes," *The Philosophical Review*, vol. 68 (1959), pp. 146-147; C. B. Martin Max Deutscher "Remembering," *ibid.*, vol. 75 (1966), p. 190; and H. P. Grice "The Causal Theory of Perception," *Proceedings of the Aristotelian Society*, Supplementary Volume XXXV (1961), pp. 143-144.

⁵ *Posterior Analytics*, *passim* esp. 75b31; 90b23; 99a22; 71b29-72a5; *Nicomachean Ethics* VI, 1139b14-36; 1140b31-1141a8.

status inferior to that of studies which are part philosophical, part scientific.

(iv) A distinction much used in recent philosophy is that between logical and non-logical necessity. Some philosophers have argued that the supposed distinction is illusory, and the most renowned argument to this effect was Quine's. Quine's view, however, had a comparatively small following in Oxford, and one of the most prominent attacks on it was launched from there.⁶

Unlike so many recent philosophers, Aristotle does not pay attention to the distinction. It looks as if he is no more interested in establishing logical necessities than in establishing non-logical ones. And it even looks as if he does not think of logical necessity as a distinct kind of necessity.⁷ (This is not to make the stronger claim that he couldn't make a statement which entailed or was equivalent to a statement of logical necessity. He very likely could, when using related concepts such as implying, entailing, and following, or incompatibility, inconsistency, and self-contradiction.)

Aristotle's attitude concerning necessity can perhaps be made intelligible by turning attention to the notion of logical possibility. Aristotle may well make claims that are equivalent to claims of logical possibility, with the aid of such concepts as implication or inconsistency. But it is one thing to make such claims, and another thing to think it appropriate to use the notion of possibility in such cases. It is often said that a mere logical possibility isn't a *genuine* possibility. And there is some reason to think that Aristotle does not view it as a possibility at all. This attitude toward logical possibility should be easy to appreciate. But now one who takes such an attitude to logical possibility has a motive for taking a corresponding attitude to logical necessity, and for refusing to regard it as a distinct kind of necessity. The penalty, if he does regard it as a distinct kind of necessity, is that he can't then correlate this kind of necessity with possibility. For there will be situations, e.g., a man's dying before 500 years of age, which enjoy no kind of possibility of being otherwise (only the misnamed logical possibility), but which nonetheless are regarded as lacking a kind of necessity (logical necessity).

(v) The last contrast to which I shall draw attention is this. Some of Aristotle's key concepts do not

mesh very easily with ours. In particular, his concepts of essence, of nature, and of separability in thought, do not. The essence of a thing is not always something that enters into the concept, as can be seen from examples. The essence of thunder is said to be the cause of the sound (*Metaphysics* 1041a20-32) which in Aristotle's time was not known, and the essence of lunar eclipse is said to be the cause of the loss of light (*Metaphysics* 1044b12-15) which at that time was still being debated. Again the essence of color is said to be its ability to stimulate the light in the surrounding medium (*On the Soul* 419a9-11). *Posterior Analytics* 93b29-94a19 suggests Aristotle may have recognized that the specification of the essence does not enter into what is signified by the words for thunder and lunar eclipse. One could say that Aristotle's essence is in some respects closer to what Locke called real essence, whereas much recent philosophy has been concerned with something more like nominal essence. One could even say that Locke, an Oxford man, contributed to this difference of interest, by treating real essence as undiscoverable, and concentrating on the nominal. It is a mistake to use the blanket notion of essentialism in describing Aristotle's position. One should rather examine what kind of essence he is concerned with.

A second example in which our concepts do not fit perfectly with Aristotle's is provided by his concept of an impossibility of nature. This seems at first sight to come very close to our concept of a logical impossibility. The most striking account of it comes in *On the Motion of Animals* 699b18-21. There Aristotle contrasts the impossibility of seeing the men on the moon, if there are any, and the impossibility of seeing a sound. This might well be taken for an illustration of the difference between logical and non-logical impossibility. But one's impression that it is such is dispelled when one turns to Aristotle's characterization of the distinction. It is characterized as being between what is of a nature to be seen, though it won't be, and what is not of a nature to be seen.⁸ Now by the nature of a thing he means the principle of change within it (e.g., *Physics* 192b20-23; *On the Heavens* 268b16; *Metaphysics* 1015a13-19). In the case of a man this would consist of the powers which constitute the human soul. And Aristotle would very likely think that these powers would rule out a man's living to be 500 years of age.

⁶ W. V. O. Quine, "Two Dogmas of Empiricism," *The Philosophical Review*, vol. 60 (1951), pp. 20-43; H. P. Grice and P. F. Strawson, "In Defense of A Dogma," *ibid.*, vol. 65 (1956), pp. 141-158.

⁷ Contrary to what might be inferred from T. Waitz, *Aristotelis Organon* (Lipsiae, 1844), vol. 1, p. 376; F. M. Cornford, *The Laws of Motion in Ancient Thought* (Cambridge, 1931), p. 21; G. E. R. Lloyd, *Polarity and Analogy* (Cambridge, 1966), p. 423.

⁸ See also *Physics* 204a3-7; *On the Soul* 422a26-29.

If so, the impossibility of his living to be 500, though not a logical impossibility, will have to be classed along with the impossibility of a sound being seen, as something that is ruled out by nature. And now this class of impossibilities is seen not to coincide neatly after all with the class of logical impossibilities.

Finally, Aristotle's concept of separability in thought does not correspond with the logical possibility of separate existence. For geometrical figures are, according to Aristotle, separable in thought from the kind of matter that undergoes change. Yet he doesn't recognize any kind of possibility such that it is possible for them to exist separately from this kind of matter. On the contrary, he often says they cannot do so.⁹ (Contrast Descartes, *Meditation* VI, on the separability in thought of himself from his body. See also Berkeley, *Introduction to the Principles*, on the separability in thought of color from extension. Sect. 7 ascribes a certain view to Locke. Later Berkeley gives his own view.) There may well be a failure of correspondence in the opposite direction too. That is, it looks as if some things which enjoy a logical possibility of existing separately may nonetheless be, in Aristotle's sense, inseparable in thought.

II

I should like now to return to the first of the contrasts mentioned and to do something to explain why such a difference should exist. As far as possible, I should like to find explanations that will enable us

to sympathize with Aristotle's views on what should be included in a definition.¹⁰

(i) One incentive, of course, for packing a lot of scientific information into his definitions is to enable them to play their role in theoretical *episteme*. They must contain enough information to facilitate deductions and scientific explanations. So they must be full, and not "empty" (*On the Soul*, 402b25-403a2; 409b14-18) like the definitions of dialectic. But this is not the only reason for Aristotle's view on what should be included in a definition. There are signs he would think some of the definitions he advocates the only fully satisfactory ones, quite apart from wanting them to play a role in theoretical *episteme*.

(ii) One important influence, I think, is the nature of the concepts in which he was particularly interested. Let us take the example of thunder, whose definition Aristotle discusses. In Aristotle's time, the concept of thunder would have been somewhat indeterminate. Its causes were still a matter for debate (*Meteorology* II.9; *Posterior Analytics* 94b32-34 Cf. Aristophanes, *Clouds*, 366-412). And its accompanying features were not known, for it was not thought to be invariably accompanied by lightning (*Meteorology* II.9, 369b6), or by storm (on bolts from the blue, see Xenophon, *Hellenica* 7.1.31; Cratinus, fragment 53, ed. Kock; Homer, *Odyssey*, XX.113). In this situation of comparative ignorance, what people meant by calling something a case of thunder might be nothing more determinate than that it was a loud crash in the sky with roughly the same causes or features (whatever those might be) as

⁹ It is hard to be sure just what Aristotle's view is, but perhaps it goes as follows. Geometrical figures exist only in physical objects. They may exist potentially, or, when the geometer thinks of them, actually (*Metaphysics*, 1051a21-33; 1078a30-31), but in either case they exist in physical objects. Of course, the very same geometrical form that is in an external physical object may exist as an object of thought within the geometer. But even then it is, like other objects of thought, embedded in an image (*On the Soul* 431b2; 432a4-5), and an image, like other mental affections, has a physiological process as its material component (*On the Soul* 403a3-b19). So the geometrical form is still embedded in a physical object, even though the geometer can think of it as separate from physical objects.

Contrast, however, the interesting account by Jaakko Hintikka, "Aristotelian Infinity," *The Philosophical Review*, vol. 75 (1966), esp. pp. 208-215, who discusses a closely related context within which Aristotle would agree with Descartes that what can be thought of can be the case.

¹⁰ Explanations can be given at various levels. One could simply point out that it is Aristotle's wish to state the cause of the definiendum that introduces obscure scientific information into the definitions of thunder, lunar eclipse, consonance, squaring the rectangle, and the rightness of the angle in a semicircle. Again, it is his desire to state the material cause, or matter, of the thing defined that brings such information into the quoted definitions of anger, a saw, and light, and into the *De Sensu*'s definitions of sense-objects. Sometimes it is by stating the function, or final cause, that Aristotle introduces obscure theory, as in *On the Soul*'s definition of color. Again it can be the purpose of the particular context that guides the composition of a definition. Thus the definitions in *On Generation and Corruption* pick from among the criteria for heat and cold those which are logically and causally fundamental, because the aim there is to find the fundamental principles of generation and corruption. Meanwhile in the psychological works, Aristotle wishes to analyze the senses by reference to their objects, and so as far as possible he there defines the sense-objects by reference to something other than the senses.

Such explanations still leave certain questions unanswered, however. For example, why should a definition be thought the right place for mentioning the causes of the thing defined in the first place? In what follows, I shall seek an answer to this question among others (see (iii)-(vi) below). And I shall look for reasons that could appeal to present-day philosophers, rather than viewing the question simply in Aristotle's terms.

belonged to the agreed instances of thunder, or at any rate to most of them. In order to mark the "whatever"-clause, we might call such concepts indefinite ones.

Now there has been one strand of thought in recent philosophy according to which one should define a thing, if possible, by giving a list of fairly specific conditions, each of them logically necessary, and jointly logically sufficient, for correct application of the word. Of course it is recognized that this is not always possible. And the ancient concept of thunder seems to be a case where it is not. Many would agree that the next best course of action here is to define by reference to examples. Placed in Aristotle's situation, one could define thunder as a crash in the sky of a certain kind but indicate the kind merely by reference to examples, not by spelling out its logical boundaries. One could define it, for example, as a crash in the sky of the kind that is often accompanied by lightning. (Only some instances of thunder were accompanied by lightning, in Aristotle's view, and so this would be a definition by reference to examples.) Indefinite concepts more or less like the ancient concept of thunder have been discussed by H. P. Grice¹¹ and by Hilary Putnam, who chose the example of multiple sclerosis.¹²

Indefinite concepts are only one type of concept which it is difficult or impossible to define by giving a list of fairly specific conditions, each logically necessary and jointly logically sufficient.¹³ What is important is that these concepts crop up particularly often in science and mathematics, to which Aristotle devoted so much of his attention. Because of the frequency with which he would encounter such concepts, we should not be surprised if he never developed a strong interest in definitions that list specific logically necessary and sufficient conditions. And we should not be surprised if he doesn't always strive to give such definitions in cases where it would be possible to do so.

(iii) There is another thing to be noticed about the ancient concept of thunder. The concept contains a reference to a cause, not a specified cause, but a cause as yet unknown. Putnam, in the two articles mentioned in the paragraph before last, has

suggested that the concept of multiple sclerosis similarly contains a reference to an undiscovered cause, this time an undiscovered virus which causes the symptoms. More recently, in a talk given at Cornell University, he has suggested that there are many concepts which contain such a reference. In particular, many of the concepts of what have traditionally been regarded as natural kinds, do so. For example, the unscientific person's concept of a lemon includes the idea that there must be some causal explanation (unknown to him) why the salient features of lemons are found together in so many cases, and of why they are not found together in all. It does not, of course, include an opinion as to what the causal explanation is, for example that it is a certain D.N.A. structure. But even though these concepts do not contain a specification of the cause, if it is true that they refer to an undetermined cause, it should be understandable why someone who had discovered the cause might wish to specify it in the definition. The ancient concepts of thunder and of lunar eclipse, which Aristotle discusses, were probably both concepts of this kind; that is, concepts that referred to an undetermined cause. And, understandably, Aristotle offers a causal explanation of thunder and lunar eclipse in their definitions, undeterred by the fact that this specification did not enter into the concept, and was not known to users of the words for thunder and lunar eclipse. He did not prefer the course mentioned above of defining thunder by reference to examples.

It was often Aristotle's insistence on stating the *cause* or *explanation* of the definiendum that led him to include obscure information, not generally known by users of the words for the definienda. Obscure explanations appear not only in his definitions of thunder and lunar eclipse, but also in his definitions of consonance, squaring the rectangle, and the rightness of the angle in a semicircle. It may have been partly from reflection on concepts like thunder and lunar eclipse, where there is a reference to an undetermined cause, and where it is reasonable to specify the cause if one can, that he was led on to giving causal explanations in the definitions of other concepts as well.

(iv) Aristotle says more about why it is important

¹¹ "The Causal Theory of Perception," *op. cit.*, pp. 143-144.

¹² *Analytical Philosophy*, ed. by R. J. Butler, series 1 (Oxford, 1962), pp. 218-221, and series 2 (Oxford, 1965), pp. 4-9.

¹³ For other examples, I can refer to the literature on this subject, not only to the articles mentioned in the preceding two notes, but also to Douglas Gasking, "Clusters," *Australasian Journal of Philosophy*, vol. 38 (1960), pp. 1-36; to William Whewell and N. R. Campbell (references given by Gasking); to Max Black, *Problems of Analysis* (London, 1954), ch. 2 (reprinted with modifications from *The Philosophical Review*, vol. 61 [1952], pp. 532-550); to Putnam, "The Analytic and the Synthetic," *Minnesota Studies in the Philosophy of Science*, vol. 3, ed. by H. Feigl and G. Maxwell (Minneapolis, 1962); and to Putnam, "It Ain't Necessarily So," *The Journal of Philosophy*, vol. 59 (1962), pp. 658-671.

to specify the cause or explanation in a definition. It is because a definition is meant to give knowledge of a thing, and of what a thing is. And one knows *what* a thing is when and only when one knows *why* it is (*Posterior Analytics* II, chapters 2, 8, 10). Similarly, one knows a thing when and only when one knows its cause or explanation (*Posterior Analytics* 71b9-12; 94a20-21; *Physics* 184a12-14; 194b18-19; *Metaphysics* 983a25-26; cf. 981a24-b6; 1044b13). This last view is put forward not as a special one of Aristotle's own, but as a popular view accepted by everyone. It needs to be explained why it should have appealed to Aristotle and others. One factor is that the view is plausible in connection with concepts such as we have just been describing. For example, it is plausible that identification of the virus will give knowledge of multiple sclerosis. If such cases are as widespread as Putnam has suggested, this could help to account for the view's being accepted. And its acceptance in turn helps to explain Aristotle's preference for specifying causal explanations.

(v) There are other ways too of explaining Aristotle's preference for incorporating scientific explanations in his definitions. One advantage of specifying the cause in a definition is that it can give unity to an otherwise diverse collection. For example, there is a variety of symptoms characterizing anger, a variety of properties constituting heat, or distinguishing silver, a variety of events counting as thunder, or eclipse, and a variety of bodily parts, and of powers of the soul, characterizing each kind of animal. By specifying a single cause of the symptoms, or properties, or events, or an explanation of why the bodily parts and soul-powers go together in a particular combination, one can make the definiendum look less of a hotch-potch.¹⁴

(vi) Again, the specification of a cause can be convenient because it gives one a way of deciding on borderline cases. At certain stages of scientific

advance, it must have been hard to see what differentiated thunder or lunar eclipse from other similar sounds or losses of light. If one can find a cause of the sound or loss of light, one will provide a way of deciding on borderline cases, and so give firmer edges to the concept.

It is interesting incidentally to notice, though this in no way helps to explain Aristotle's procedure, that in recent times, with the advance of science, this kind of decision procedure, namely looking to see if a certain cause is present, has come to be used increasingly for determining whether one has a case of thunder, of eclipse, of consonance, of a certain chemical or metal, of a certain disease, of male or female sex, and so on. Moreover, the *specification* of the correct causal theory tends gradually to enter first the specialist's concept, and eventually the layman's. As a result, the concepts of scientists have come in many cases to answer closely to the pattern of definition that Aristotle prescribes.

(vii) We have noticed that Aristotle included in his definitions information that was not known to users of the words for the definienda. And we have been saying that this information often comes in because of Aristotle's interest in mentioning causes and explanations. In *Topics* VI.4, he addresses himself to a different reason for not confining definitions to well-known information. There he complains that if one defines by reference to what is known to ordinary people, one may be landed with a multiplicity of definitions, corresponding to the different states of knowledge of different people, or of the same person at different times. This good point would probably have been particularly clear to Aristotle in the case of natural kinds. It is true that different people do have very different scraps of information about gold, salt, eagles, and antelopes. And this does supply a reason for not tying one's definitions to the knowledge or ignorance of word users.

¹⁴ We might add that Aristotle attaches importance to a definiendum's having some unity, and that he thinks one function of the form or essence is to give unity to a thing (see *Metaphysics* V.6; VII.4 and 12; VIII.6; X.1).

There is a complication in the case of heat, for *Parts of Animals* 648a36-649b9 argues that one cannot give it a single definition, but must give different definitions corresponding to the different powers that characterize it. (See the similar treatment of the soul, *On the Soul* 414b20-28.) However, *On Generation and Corruption* 329b26 defines heat by reference to a single power, which is perhaps regarded as casually explaining the other powers connected with heat. (See H. H. Joachim's note on 329b26-30 in *Aristotle on Coming-to-be and Passing-away* (Oxford, 1922), and see *Meteorology* IV, if the fourth book of the *Meteorology* is by Aristotle.)

We must admit that our reference to the bodily parts of the various animal species, and the powers which constitute the soul, is rather speculative. Aristotle assiduously seeks the causes which make these go together in certain combinations and not in others. But did he hope eventually to include such causes in the definition of each animal species (see D. M. Balme, "Aristotle's Use of Differentiae in his Zoology" in *Aristote et les Problèmes de Méthode, op. cit.*), or in the definition of the soul of each kind of living thing? There is a much-discussed call for a causal definition of the soul in *On the Soul* 413a13-20. But it would be somewhat overbold to interpret it as referring to this kind of cause.

The case of anger, thunder, and lunar eclipse is clearer. Here Aristotle does include the unifying cause in his definition.

I have been arguing that one thing that could have influenced Aristotle's view about what should go into a definition is the nature of the concepts in which he was interested. His interest in the concepts of science, of mathematics, and of natural kinds, could have given him an incentive to offer scientific explanations in his definitions, and not to tie himself to what was known by users of the words for the definienda. These interests would also give him no incentive to prefer definitions that listed specific logically necessary and sufficient conditions. This should help us to understand the existence of the first contrast I listed above, the contrast concerning the question of what should go into a definition.

(viii) I have one more explanation to offer for the existence of this contrast. It is the fact that Aristotle does not recognize logical necessity as a distinct kind of necessity. In *Physics* II.9, 200b4–8, he expresses the uncontroversial idea that a definition of *X* may need to mention what is necessary for the performance of *X*'s function, and hence necessary for *X*. He cites the definition of a saw. It may have to mention that teeth of a certain sort are necessary for the performance of a saw's function, and so necessary for a saw. If this means teeth with a certain degree of hardness, sharpness, and so on, the necessity is one we could plausibly take as a logical necessity. (Butter or water cannot form the teeth of a saw.) But then he crosses over, with apparent indifference, to a necessity that was clearly not logical, the necessity of iron for the obtaining of such teeth. Iron may also need to be mentioned in the definition, so he says. There is a similar passage in *On the Soul*, 403a25–b9, which discusses the definition of anger. The preferred definition of anger will include reference to the boiling of the blood around the heart. This boiling, on Aristotle's theory, is necessary (403b3) for the existence of the form of anger, and so for the existence of anger. But there is no attempt to show that the necessity is logical.

Part of the contrast under discussion consists in the fact that Aristotle does not try to confine the definitions of saw and anger to necessities which are logical. We shall have one explanation of this, if, for reasons given earlier in the paper, Aristotle did not regard logical necessity as a distinct kind of necessity. He would then not think that iron was necessary in a different way from hardness and sharpness. And so he would lack this incentive for treating the necessity of iron in a different way from the necessity of hardness and sharpness, and excluding it from the definition.

III

The contrasts cited at the beginning of this paper have consequences for the way in which one ought to approach Aristotle. There will be restrictions on how one ought to describe his views. Can one, for example, legitimately describe him as believing such-and-such to be a matter of logical necessity? And there will be restrictions on what inferences one can draw about his views, when he talks of nature, essence, definition, or separability in thought. It would be easy, but needlessly polemical, to cite commentators who neglect these restrictions.

We can illustrate the point by reference to the *Nicomachean Ethics*. I started the paper by saying that the contrasts I had in mind were not very noticeable there. Let us finish by saying that nonetheless the *Ethics* does not escape the consequences of these contrasts. I once puzzled, as an undergraduate, over whether it is a logical or a non-logical necessity that Aristotle postulates at *Nicomachean Ethics* 1147a27. He is there speaking of the practical syllogism, and he states certain circumstances under which it is *necessary* for a man to act. One consequence of our discussion above is that it may be misleading to give either answer to this question. It will be misleading if it suggests to anyone that Aristotle recognized logical necessity as a distinct kind of necessity. My present view on this passage is that Aristotle regards the necessity in question as a matter of definition. But it is a further consequence of our discussion that since he is willing to include so much in a definition, it doesn't follow immediately that he regards the necessity as a conceptual one. The defense of my interpretation of the passage I shall postpone to another occasion.

APPENDIX

It might be useful by way of an appendix to cite some of the passages that are relevant to whether Aristotle recognized logical necessity as a distinct kind of necessity. It is tempting to suppose that he did so, in cases where (a) he claims that some necessity exists, and we can see that there is a logical necessity, or that only a logical necessity could reasonably be argued for. (b) He gives his grounds for a claim of necessity, and we can see that they would support a claim of logical necessity. (c) He makes a statement (involving, say, the notion of apodeictic proof, or of contraries, or the notion of being a prerequisite of thought), and we can see that it entails a statement of logical necessity. (d) He

uses concepts such as those of implying, entailing, and following, or of incompatibility, inconsistency, and self-contradiction, in such a way that we can see that what he says is equivalent to a claim of logical necessity. Such passages do not, however, show that he regarded logical necessity as a necessity of a distinct kind.

It is easy to be misled also by a large number of concepts which he did possess, and which at first sight seem to correspond closely to our logical necessity, or to members of the same immediate family, logical possibility and impossibility. We have already encountered two such concepts, namely that of an impossibility of nature, and that of separability in thought.

Let us now give some evidence on the other side. One kind of evidence is provided by Aristotle's habit of jumbling together logical and non-logical necessities, with apparent indifference, in cases where we would want to separate them. It is a habit that shows up in many contexts. It will suffice here to list some of the references without discussing them in detail.

- (a) The jumbling of examples is to be found in his illustrations of the notion of hypothetical necessity, throughout *Physics* II.9.
- (b) It crops up again when he illustrates the notion of necessary signs (*Rhetoric* 1375a22–b21. Note that fever is a kind of illness, *Topics* 123b35–36).
- (c) He has a principle that, in connection with certain classes of entity, what is always so is necessarily so. (References are given in Jaakko Hintikka's five articles on the principle of plenitude to be cited below. The present principle is a corollary of the principle of plenitude.) No warning is offered that the necessity of what is always so will in some cases be logical, in some cases not.
- (d) The premisses of the syllogisms that enter into theoretical *episteme* include definitions. These definitions tell us of properties that attach essentially (*kath' hauta*) to their subjects. In *Posterior Analytics* I.4 and 6, Aristotle says that the premisses are necessary (73a24; 74b5–6; 74b13–75a17), and that properties which attach essentially to their subjects attach necessarily to them (73b16–24; 74b7; 75a28–29). He does not warn us that the necessity is not always logical, that, for example, the correct causal explanation which constitutes the essence of thunder or lunar eclipse

did not attach by logical necessity to these phenomena.

- (e) The notion of necessity in demonstrations, i.e., in the syllogisms appropriate to theoretical *episteme*, appears in *Metaphysics* 1015b6–8 and 1064b33–34. It emerges from 1015b8–9 that it is not a necessity linking premisses with conclusion, but the necessity which is transferred to the conclusion thanks to the premisses themselves expressing necessary truths. So it is the same kind of necessity as is described in (d) above, and again no warning is given that it includes logical and non-logical necessities.
- (f) *Metaphysics* V.5 offers a comprehensive list of the different uses of the word for necessity. (The vague expression "different uses" is meant to cover both Aristotle's talk of "homonymy," and his talk of "being said in different ways.") It is not quite certain whether we should treat as a significant omission Aristotle's failure to devote a separate listing to logical necessity. Does it show that he didn't think of this as a distinct kind of necessity? Or are we only to understand that, by his criteria for difference of use, "necessity" does not have different uses in the phrases "logical necessity" and "non-logical necessity," and that this does not preclude there being a difference in kind? To decide, one must look at his criteria for difference of use in *Categories*, chapter 1, and *Topics* I.15, and at his examples of difference of use throughout *Metaphysics* V.¹⁵ Even if we do not treat the omission as evidence, the chapter is still relevant to our subject, for it provides another example of the jumbling phenomenon. Aristotle does not give a warning that examples of logical and of non-logical necessity will fall under the heading of hypothetical necessity, and again under the heading of what is incapable of being otherwise (a heading illustrated by the necessity in demonstrations, discussed in (e) above).
- (g) In the famous discussion of future contingents in *De Interpretatione*, chapter 9, some of the necessities mentioned are logical, some non-logical. Some kinds of necessity (e.g., the necessity of what always happens in one way) will cover both logical and non-logical examples. And in some cases the status of the necessity mentioned is hard to determine.
- (h) Finally, there is a rather different kind of evidence. Aristotle held a view, sometimes called the principle of plenitude, according to which, in connection with certain classes of entity, what is possible must at some time be actual.¹⁶ The view

¹⁵ See also Jaakko Hintikka, "Aristotle and the Ambiguity of Ambiguity," *Inquiry*, vol. 2 (1959), p. 137.

¹⁶ The view is documented by Jaakko Hintikka in five articles; "Necessity, Universality and Time in Aristotle," *Ajatus*, vol. 20 (1957), pp. 65–90; "Aristotle and the 'Master Argument' of Diodorus," *American Philosophical Quarterly*, vol. 1 (1964), pp. 101–114; "The Once and Future Sea Fight," *The Philosophical Review*, vol. 73 (1964), pp. 461–492; "Aristotelian Infinity," *ibid.*, vol. 75 (1966), pp. 197–218; and "A. O. Lovejoy on Plenitude in Aristotle," *Ajatus*, vol. 29 (1967), pp. 5–11.

is not obviously false in connection with certain non-logical possibilities. But it is strikingly incorrect in connection with what is merely logically possible. So one needs to explain why he fails to point out that the principle does not apply to what is possible only in this way. He does point out areas in which the principle fails to hold, but none of these caveats looks like an effort to exempt what is merely logically possible. The simplest explanation

would be that he does not recognize what is possible only in this way as possible at all. Now Aristotle is very much alive to the close connections between the possible and the necessary. So he is likely to treat the necessary in a corresponding way. And the corresponding way of treating the necessary, so it was argued above, is not to regard it as a distinct *kind* of necessity, which something may lack when it has other kinds.¹⁷

Cornell University

Received May 17, 1968

¹⁷ This represents a summary of results of some work in progress. For reasons of space, I have had to leave a number of my claims about Aristotle undefended here.

I have received generous help from Robert M. Adams, Philip De Lacy, Keith Donnellan, Robert Ennis, Bruce Goldberg, Charles Kahn, and Sydney Shoemaker.

The references to Oxford are not meant to suggest that there is only one approach to philosophy in Oxford. It is merely that the tendencies in recent philosophy with which I wish to contrast Aristotle's work have been well represented there, while the most prominent examples of rival tendencies have come from philosophers elsewhere, such as Quine and Carnap.

V. HOW TO BURY THE MIND-BODY PROBLEM

RICHARD TAYLOR

THE mind-body problem, in all its variants, is a philosophical fabrication resting on no genuine data at all. It has arisen from certain presuppositions about matter and human nature familiar to philosophy from the time of the Pythagoreans, presuppositions which have persisted just to the extent that they have been left unexamined. And they have not been questioned very much simply because they are so familiar.

There are vexing, unsolved problems of psychology and problems of mental health, but there are no mind-body problems. And there are problems of "philosophical psychology," as they are sometimes called today—problems of perception, sensation, the analysis of deliberation, of purposeful behavior, and so on—but there are no mind-body problems.

The reason why there are no mind-body problems is the most straightforward imaginable: It is because there are no such things as *minds* in the first place. There being no minds, there are in strictness no mental states or events; there are only certain familiar states, capacities, and abilities which are conventionally but misleadingly called "mental." They are so-called, partly in deference to certain philosophical presuppositions, and partly as a reflection of our lack of understanding of them, that is of our ignorance.

Men and women are not minds, nor do they "have" minds. It is not merely that they do not "have" minds the way they have arms and legs; they do not have minds in any proper sense at all. And just as no man or woman has or ever has had any mind, so also are cats, dogs, frogs, vegetables, and the rest of living creation without minds—though philosophers of the highest rank, such as Aristotle, have felt driven to say that all living things, vegetables included, must have souls (else how could they be *living* things?) just as others of similar eminence, like Descartes, have thought that men must have minds, else how could they be *thinking* things? Today, when philosophers talk about mind-body problems, and advance various claims concerning the possible relationships between "mental" and "physical" states and events, they are, of course, talking about men. But they might as well be talking about frogs, because the presuppositions that give

rise to these theories apply to other animals as well as to men.

I. PHILOSOPHICAL ARGUMENTS FOR THE EXISTENCE OR NONEXISTENCE OF THINGS

There cannot be any philosophical argument proving that something does or does not exist, so long as the description or definition of it is self-consistent. Thus there cannot be a philosophical argument proving that men do or do not, as some medieval thinkers believed, have an indestructible bone in their bodies. One can only say that such a bone has never been found (which is not a philosophical argument) and then exhibit the groundlessness or falsity of the presuppositions that gave rise to the belief in the first place. (In this case it was certain presuppositions concerning the requirements of the resurrection of the body.) Similarly, there can be no philosophical argument proving that men do or do not have souls, spirits, or minds, or that there are not *sui generis* mental states or events, assuming that these can be described in a self-consistent way. One can only note that such things have never been found in any man, living or dead, and then exhibit the arbitrariness and apparent falsity of the presuppositions that give rise to these opinions in the first place. Now of course, as far as *finding* them goes, many philosophers claim to find them all the time, *within themselves*. They are alleged to be *private* things, deeply hidden, discernible only by their possessors. All they really "find," however, are the most commonplace facts about themselves that are perfectly well known to anyone who knows anything at all—but of this, more later.

II. THE GRAND PRESUPPOSITION OF THE MIND-BODY PROBLEM

What I must do now, then, is consider the presupposition that has given birth to the so-called "mind-body" problem, and show that there is nothing in it at all that anyone needs to believe; that, on the contrary, we have good evidence that it is false.

The presupposition can be tersely expressed by saying: *Matter cannot think*. That is the way a

Cartesian would put it, but philosophers now spell it out a little better. Thus, we are apt to be told that thinking, choosing, deliberating, reasoning, perceiving, and even feeling, are not concepts of physics and chemistry, so that these terms have no application to bodies. Since, however, men do think, choose, deliberate, reason, perceive and feel, it follows that men are not "mere bodies." They are instead minds or souls or, as it is more common to say today "selves" or "persons," and such terms as "is thinking," "is choosing," "is perceiving," etc., are not physical or bodily but *personal* predications. A man may be in one clear sense a physical object, having arms and legs and so on, but a person is not just that visible and palpable object; there is more to a self or person than this. For it is the self or person that thinks, chooses, deliberates, feels, and so on, and not his body or some part of it.

Again—and this is really only another way of expressing the same presupposition—we are apt to be told that thoughts, choices, reasons, feelings, etc., are not physical things. It makes no sense to ask how large a thought is, whether it is soluble in alcohol, and so on. Yet these things do exist—any man can be aware of them, "within himself." Hence, that "self" within which such things occur must be something more than or other than the body. It might be just the totality of all those nonphysical ("mental") things, but in any case it is mental in nature, so a self or person is not the same thing as his body.

Or again, in case one boggles at calling thoughts, feelings, and the like, "things," at least (it is said) no one can deny that they are events or states. But they are not events or states that occur or obtain in the laboratories of physicists and chemists—except in the sense that they sometimes occur in physicists and chemists themselves, who sometimes happen to be in laboratories. No one could ever truly represent whatever might be happening in a test tube or vacuum tube as the transpiring of a thought or feeling. These things just do not—indeed, obviously could not—happen in test tubes or vacuum tubes, because they are not the *kind* of event involving changes of matter. They are a kind of "mental" event. And since these things do, obviously, happen in men, then things happen in men which are non-physical, "mental," in nature. And so on.

III. "SELVES" OR "PERSONS" AS MINDS AND BODIES

The word "self" and the plural "selves" are fairly common items of contemporary philosophical

vocabulary. These words never occur outside of philosophy, except as suffixes to personal pronouns, but in philosophical contexts they are sometimes taken to denote rather extraordinary things. Selves are, indeed, about the strangest inhabitants of nature that one can imagine—except that, as sometimes described in philosophy, they are not even imaginable in the first place, being quite non-physical. You cannot poke a self with a stick; the nearest you can come to that is to poke his body. The self that has that body is not supposed to be quite the same thing as his body—that is a (mere) physical object, a possible subject matter for physics and chemistry. *That* is not what thinks, reasons, deliberates, and so on; it is the self that does things like this.

At the same time, selves are never doubted to be the same things as *persons*, and persons are thought to be the same things as people, as men. And there is no doubt at all that men are visible, palpable objects, having arms and legs and so on: That they are, in short, physical objects. So the thing becomes highly ambiguous. We do not, in contexts in which it would seem silly or embarrassing to do so, have to say that selves (men) are spirit beings (minds) which in some sense or other happen to "have" bodies. Clearly men are visible and palpable things, that is, are bodies. We can say that all right. But at the same time we need not say—indeed, *must* not say—that men are just (mere) bodies. There is, after all, a difference between a man's body, and that which thinks, perceives, feels, deliberates, and so on; and those are things that men (selves) do, not things that bodies do. Or again, there is, after all, a difference between bodily predicates (weighs 160 pounds, falls, is warm, etc.) and personal predicates (chooses, believes, loves his country, etc.). The former can be predicated of a man's body, just like any other body, but it would "make no sense" to predicate the latter of any (mere) body, and hence of any man's body. They are only predicated of persons. So even though selves are persons and persons are men and men are visible, palpable beings, we must not think that they are just nothing but physical beings. They are physical bodies with minds, or, as some would prefer, minds with physical bodies or, as most writers on this subject want to say, they are somehow *both*.

So the "mental" is discriminated from the (merely) "physical," and the mind-body problem emerges at once: What is the *connection* between them? What is the relationship between men's minds and their bodies? Or between mental and

physical events? Or between personal and physical predicates? Anyone who raises this question—for these all amount to one and the same question—can see at once that it is going to be extremely difficult to answer. And this means that it is capable of nourishing a vast amount of philosophy. It has, in fact, kept philosophers on scattered continents busy for hundreds of years, and even today claims much of the time of philosophical faculties and their proteges. It seems a conceit to undertake to put an end to all this, but that is what I propose now to do.

IV. MENTALISM AND MATERIALISM

Consider the following two theses:

- (I) A person is not something that has, possesses, utilizes, or contains a mind. That is, a person is not one thing and his mind another thing. A person or self and his mind are one and the same thing.
- (II) A person is not something that has, possesses, utilizes, or occupies a body. That is, a person is not one thing and his body another thing. A person or self and his body are one and the same thing.

We can call these two theses "mentalism" and "materialism" respectively, since the first asserts that men are minds and not bodies, and the second that they are bodies and not minds.

Now the first thing to note about these two rather crudely stated theses is that both of them cannot be true, since each asserts what the other denies. They could, of course, both be false, since a person might be identical neither with his body nor with his mind (though it is hard to think of any other candidate for the title of "person"), or a person might somehow be identical with the two of them at once. These two simple theses are, nevertheless, a good starting point for discussion, and I am going to maintain that (II), the materialist thesis, is absolutely true.

Philosophers have tended to regard (I), or some more sophisticated version of it, as correct, and to dismiss (II) as unworthy of consideration. In fact, however—and it is hard to see how this could have been so generally overlooked—any philosophical argument in favor of (I) against (II) is just as good an argument for (II) against (I). This I shall illustrate shortly.

In the meantime, let us give what is due to the humble fact that there are considerations drawn from common sense, indeed from the common knowledge of mankind, which favor, without

proving (II). It is common knowledge that there are such things as human bodies, that there are men and women in the world. There is also one such body which everyone customarily, and without the least suggestion of absurdity, refers to as himself; he sees himself in the mirror, dresses himself, scratches himself, and so on. This is known, absolutely as well as anything can be known, and if any man were to profess doubt about it—if he doubted, for example, that there are such physical objects in the world as men and women, and therefore doubted the reality of his own body—then that man would have to be considered *totally* ignorant. For there is nothing more obvious than this. A man would be ignorant indeed if he did not know that there are such things as the sun, moon, earth, rivers, and lakes. I have never met anyone so ignorant as that. But a man who did not even know that there are men and women in the world, and that he—his body—was one of them, would be totally ignorant.

Now there is no such common knowledge of the existence of minds or souls. No one has ever found such a thing anywhere. Belief in such things rests either on religious persuasion or on philosophical arguments, sometimes on nothing but the connotations of familiar words. Such beliefs are opinions, easily doubted, and nothing that anyone knows. If a man denies that such things exist, as many have, then he exhibits no ignorance; he expresses only scepticism or doubt concerning certain religious or philosophical presuppositions or arguments.

If, accordingly, we are seeking some sort of thing with which to identify persons, then this is a *prima facie* consideration in favor of identifying them with their bodies, with things we know to be real, rather than with things postulated to suit the requirements of philosophical arguments or religious faith. This does not prove that men are nothing but bodies, of course, but it is enough to show that, since we know there are such things as persons, and we know there are such things as men (living human bodies), we had better regard these as the very same things *unless* there are some facts which would prohibit our doing so. And I shall maintain that there are no such facts. There are only philosophical arguments, not one of which proves anything.

The arguments for mentalism. I shall now consider the arguments I know, already adumbrated, in favor of what I have called mentalism. Of course not all philosophers who take seriously the mind-body problem subscribe to this simple thesis as I have formulated it, but the more sophisticated versions can be considered as we go along, and it will be

seen that the arguments for these are equally inconclusive.

The first argument. There are certain predicates that undoubtedly apply to persons, but not to their bodies. Persons and their bodies cannot, therefore, be the same. One can sometimes truly say of a person, for example, that he is intelligent, sentimental, that he loves his country, believes in God, holds strange theories on the doctrine of universals, and so on. But it would sound very odd—indeed, not even make sense—to assert any such things of any physical object whatever and hence of any man's body. It would at best be a confusion of categories to say that a certain man's *body* loves its country, for example.

Reply. If the foregoing is considered a good argument for the non-identity of persons and bodies, then the following is obviously just as good an argument for not identifying them with their minds: There are certain predicates that undoubtedly apply to persons, but not to their minds. A person and his mind cannot, therefore, be the same. One can sometimes truly say of a person, for example, that he is walking, ran into a post, is feverish, or that he fell down. But it would sound very odd—indeed not even make sense—to assert such things of any mind whatever. It would at best be a confusion of categories to say, for instance, that a certain man's *mind* ran into a post.

Considerations such as these have led many philosophers to affirm that a person or the "true self" is neither a mind, nor a body. Hence, a person must be (a) something else altogether or, as some would prefer to say, the term "person" must express a "primitive" concept or (b) both mind and body; i.e., a person must be something having both mental and physical properties.

The former of these alternatives is simply evasive. Persons are real beings, so there must be existing things which are persons. If when we bump into a man we are not bumping into a person, and if at the same time we are not referring to a person when we say of someone that he is thinking, then it is quite impossible to see what is left to fill the role of a person. The word "person" may indeed be a primitive one, but this, I think, only means that such arguments as the two just cited are equally good and equally bad.

The second alternative that persons are beings having both mental and physical properties, is obviously only as good as the claim that there are such things as "mental properties" to begin with. Indeed, it is not even that good, for just as a physical

property can be nothing but a property of a physical thing, i.e., a body, so also a mental property can be nothing but the property of a mental thing, i.e., a mind. For something to count as a physical property of something it is sufficient, and necessary, that the thing in question is a physical object. By the same token, for something to count as a mental property it is sufficient, and necessary, that it be the property that some mind possesses. Any property whatsoever that can be truly claimed to be the property of some body, animate or inanimate, is a physical property; the assertion that some body possesses a nonphysical property is simply a contradiction. This second alternative, that persons are beings possessing both physical and mental properties, therefore amount to saying that a person is at one and the same time *two* utterly different things—a body with its physical properties and a mind with its mental properties. These are not supposed to be two things in the same sense that a family, for instance, is a plurality of beings consisting of husband, wife, and perhaps one or more children, but two wholly disparate kinds of beings having, as Descartes put it, nothing in common. Now this is no resolution of the antithesis between what I have called mentalism and materialism. It is only a reformulation of that issue. For now we can surely ask: Which of these two is the person, the true self? The body which has a mind, or the mind which has a body? And we are then back where we started.

The second argument. This argument consists of pointing out the rather remarkable things that a person can do but which, it is alleged, no physical object, of whatever complexity, can do, from which it of course follows that a person is not a physical object and hence not identical with his own body. A person, for example, can reason, deliberate about ends and means, plan for the future, draw inferences from evidence, speculate, and so on. No physical objects do such things, and even complicated machines can at best only simulate these activities. Indeed, it would not even make sense to say that a man's body was, for example, speculating on the outcome of an election, though this would not be an absurd description of some person. A person, therefore, is not the same thing as his body, and can only be described in terms of certain concepts of mind.

Reply. This argument is not very different from the first; it only substitutes activities for properties which are baptized "mental." And one reply to it is the same as to the first argument; namely, that since persons often do things that no mind could do—for instance, they run races, go fishing, raise families,

and so on—then it follows that persons are not minds.

A far better reply, however, and one that is not so question-begging as it looks, is to note that since men do reason, deliberate, plan, speculate, draw inferences, run races, go fishing, raise families, and so on, and since the men that do all such things are the visible, palpable beings that we see around us all the time, then it follows that *some* physical objects—namely, men—do all these things. All are, accordingly, the activities of physical objects; they are not activities divided between a physical object, the visible man, on the one hand, and some invisible thing, his mind, on the other.

Consider the statement: "I saw George yesterday; he was trying to figure out the best way to get from Albany to Montpelier." Now this statement obviously refers, in a normal context, to a person, and it is perfectly clear that the name "George" and the pronoun "he" refer to *one and the same* being, that person. And what they both refer to is something that was seen, a certain man's body; they do not refer to some unseen thing, of which that body is some sort of visible manifestation. If that were so, then the statement would not really be true. And in any case, it would be embarrassingly silly to suppose that a more accurate rendition of the thought expressed in this statement might be: "I saw George's body yesterday. His mind was trying to figure out how to get (how to get what?) from Albany to Montpelier." It is, accordingly, one and the same thing which (a) is seen, and (b) figures and plans, and that thing is undoubtedly the physical object George. Now if conventions incline us to describe figuring out something as a "mental" activity, then we shall have to say that some purely physical objects—namely, living men—engage in mental activities. But this is simply misleading, if not contradictory, for it suggests that we are ascribing to a physical object an activity of something that is not physical, but mental. It would, therefore, be far better to say that some physical objects, namely, men or persons, sometimes perform physical activities such as figuring and planning which are quite unlike those we are accustomed to finding in certain other physical objects such as machines and the like.

The third argument. This argument, the commonest of all, is to the effect that while there may or may not be such things as "minds" (whatever that might mean), there are indisputably certain nonphysical things which are quite properly called "mental," as anyone can verify within himself. Indeed, it is sometimes claimed that nothing, not even the reality of

our own bodies, is as certain as the existence of these mental things, which are perceived "directly."

Reply. What are here referred to as mental entities are, of course, such things as thoughts, mental images, after-images, sensations, feelings, and so on. Pains are frequently mentioned in this context, being, presumably, things whose existence no one would question. Having got to this point then the next step, of course, is to speculate on the connection between these mental things and certain "physical" states of the body. They evidently are not the same, and yet it is hard to see what the connection could be. Speculation also extends to such questions as whether two or more men might have "the same" pain, or why it is impossible that they should in view of the fact that they can hold common possession of ordinary "physical" things like clocks and books. Again, curiosity is aroused by the fact that a mental image, for instance, seems to have color, and yet it somehow can be perceived only by one person, its owner. Again, images sometimes seem to have shape—enough so that a perceiver can distinguish one from another, for instance—and yet no assignable size. Here, really, is a gold mine for philosophical speculation, and such speculations have filled, as they still fill, volumes.

Now surely there is a *better* way to express all that is known to be true in all this, and it is a way that does not even permit these odd theories to get started. What we know is true, and all we know is true, is that men think, sense, imagine, feel, etc. It is sheer redundancy to say that men think things called "thoughts," sense things called "sensations," imagine "images," and feel "feelings." There are no such things. And to say there are no such things is *not* to deny that men think, sense, imagine, and feel.

What, for instance, does it mean to say a man feels a pain in his foot? Absolutely nothing, except that his foot hurts. But this hurting, what sort of thing is it? It is not a thing at all; not a thing felt, and certainly not a mental thing that is felt *in his foot*. It is a state, and in no sense a state of his mind, but a straightforward state of his foot. But can that be a *physical* state? Well, it is assuredly a state of his foot, and that is a physical object; there is nothing else—no spirit foot, no spirit being, no spirit mind—that it can be a state of. Why, then, cannot other people have that same state? Why cannot other people feel the same pain I feel in my foot? And if it is a physical state, why cannot we open the foot and *see* it there? Or make some straightforward test of its presence in another man's foot?

To ask questions like these is just not to understand

what is meant by describing an object as being in a certain state. Consider a piece of molten lead. Now this molten state, what sort of thing is it? The answer is that it is not a thing at all; it is a state or condition of a thing. Is it a physical state? Well, it is a state of the lead, and that is a physical object; there is nothing else for it to be a state of. Why, then, cannot another piece of lead have that same state? Why cannot something else have the molten state of this piece of lead? Of course something else can, in the only meaningful sense that can be attached to such a question; that is, another piece of lead, or some things which are not lead can melt the same way this piece of lead melted. To ask why another piece of lead cannot have the molten state of this piece of lead is, of course, unintelligible, unless it is interpreted the way just suggested, in which case the answer is that it can. But similarly, to ask why another man cannot have the pain that this man is feeling is also unintelligible, unless construed as the question why other men cannot suffer pain, in which case its presupposition is wrong—they can. And if the piece of lead's being melted is a "physical" state, why can we not separate the lead into drops and see that state? Simply because it is a state of the lead, and not some other thing contained in the lead. Indeed, to separate it into drops is to see, not its meltedness (there is no such thing), but that it is melted—that is just the test. We do not have to ask the lead whether it is melted, and rely upon its testimony; we can tell by its behavior. And in the same way we can sometimes—admittedly not always—see that a man is suffering, without having to ask him. That we sometimes go wrong here does not result from the fact that his suffering is something quite hidden within him, which he alone can find and then report; there is nothing hidden, and nothing for him to find. Still, there is a straightforward way of testing whether a piece of lead is melted, and there is no similarly straightforward way of testing whether a man's foot hurts—he may only be pretending it does. Does this indicate that there might be a pain, which he has found in his foot but might conceal, as he might conceal the contents of his wallet? Surely not; it shows only that men, unlike pieces of lead, are capable of dissimulating. No philosophy was needed to unearth that commonplace fact. It is easier to test for the presence of some states of properties than others, and this is true not only of the states of men's bodies, but of everything under the sun. But things that are hard to establish do not, just by virtue of that, warrant the title of "mental."

Similar remarks can be made about images, which are frequent candidates for the role of mental entities. When queried about their mental imagery, people often will describe it in colorful detail and even with pride, not unlike the regard one might have for a precious gem accessible only to himself. It turns out, though, that all one thereby describes is his power of imagination, which is, of course, sometimes quite great. To say that one has a lively imagination, even great powers of imagination, does not mean that he can create within his mind, *ex nihilo*, things called "images" and composed of some mental, nonphysical, spiritual material. There is no material that is non-material, and there are no images composed of this or anything else—except, of course, those physical objects (pictures, etc.) visible to anyone who can see, which are rightly called images of things. When someone *sees* something, there is (i) the man who sees, and (ii) the thing seen; for instance, some building or scene. There is not, between these, a third thing called the appearance of what is seen; philosophers are pretty much agreed on this. But similarly, when someone *imagines* something or, as it is misleadingly put, "forms an image" of it, there is (i) the man who imagines, and (ii) sometimes, but not always, something that he imagines; for instance, some building or scene, which might or might not be real. There is not, between these, a third thing called the image of what is imagined. There is just the imagining of the thing in question. And to say that a man is imagining something is to say what he is doing, or perhaps to refer to some state he is in; it is not to refer to some inner thing that he creates and, while it lasts, exclusively possesses.

It is enough, it seems to me, to point this out; that is, to point out that we can say all we want to say about men's powers of imagination without ever introducing the substantive "an image." Philosophy is robbed of nothing by the disposal of these, and there is absolutely no fact about human nature which requires us to affirm their existence. But if one does insist upon the reality of mental images, and professes, for instance, to find them right in his own mind by introspecting—and it is astonishing how eager students of philosophy seem to be to make this claim—then we can ask some very embarrassing questions. Suppose, for instance, one professes to be able to form a very clear image of, say, the campus library—he can bring it before his mind, hold it there, perhaps even turn it bottom side up, and banish it at will. We ask him, then, to hold it before his mind and count the number of steps in the

image, the number of windows, the number and disposition of pigeons on the roof, and so on. He could do these things if he had a photograph of the thing before him. But he cannot do them with the image, in spite of the fact that it is supposed to be right there "before his mind," easily and "directly" inspectable. He can tell how many steps there are only if he has sometime counted the steps on the building itself (or in a photograph of it) and now *remembers*—but that is not counting the steps in the image. Or he can *imagine* that it has, say, 30 steps, and then *say* "30"—but that is not counting anything either; it is only a performance. The image he professes to "have" there, so clearly and with such detail, does not even exist. He claims to have produced in his mind an image of the library; but all he has actually done is imagine the library.

What, then, is imagining something? Is it an activity, a state, or what? It does not really matter here how we answer that; it is only *not* the producing of an entity called a "mental image." Let us suppose for this context, then, that to be imagining something is to be in a certain *state*. Is it, then, a *physical* state? Well, it is a state of a man, just as drunkenness, sleep, perspiration, obesity, etc., are sometimes states of this man or that. What is meant by asking whether these are "physical" states, other than asking whether they are states of a physical object? What shall we say of being in a state of sleep, for instance? It is the state of a man, and a man is a physical—that is, a visible and palpable—being. You cannot poke a man's state of imagining something with a stick; all you can do is poke him. That is true. But you cannot poke his somnolence with a stick either. There is nothing to poke; there is only the man sleeping, or the man imagining, or the man becoming drunk, or whatever.

How then can a man, if he is nothing but a (mere) physical object, be in such a state as this, that is, of imagining something? If he is only a body and can do this, why cannot sticks and stones be in such a state, for are they not bodies too? The answer is: For just the same reason that sticks and stones cannot be drunken, asleep, perspiring, obese, or hungry; namely, that they are sticks and stones and not men. The reason is not that they lack minds. Even if they had them, they still could not be drunken, asleep, perspiring, obese or hungry, for they would still be sticks and stones and not men.

The fourth (and last) argument. It is fairly common for people, including philosophers, to say that they can perfectly well imagine surviving the death of their bodies, which would be quite impossible for

anyone who supposed that he and his body were one and the same thing. Admittedly no one knows whether there is any survival of death, but it is at least not necessarily false. The doctrine of metempsychosis, for example, though there may be no reason for believing it, cannot be shown to be impossible just on philosophical grounds. It would be impossible, however, if a person and his body were identical, and so would any other form of survival. We know the fate of the body: dust. If I am the same as my body, then it is logically impossible that I should not share that fate.

Reply. All this argument shows is that not everyone, perhaps even no one, *knows* that he and his body are one and the same thing. It does not in the least show that, in fact, they are not. Some things, like the Evening Star and the Morning Star, which some are accustomed to thinking of and describing as different things, nevertheless do turn out to be the same.

Suppose a god were to promise me a life after death—promising, perhaps, to have me (the very person that I am) reborn elsewhere with a different body. Now such a promise might quicken a real hope in me, provided I am capable (as everyone is) of thinking of myself as being something different from my body. But the fact that I can think such a distinction does not show that there is one, and in case there is not—in case I happen to be identical with my body—then of course no god could fulfill such a promise. Consider this analogy: If an enemy of our country did not know that Albany is (the same thing as) the capital of New York, then he might be very interested in a proposal to bomb the one but to spare the other. It would nevertheless be a proposal that no one could carry out. The fact that someone who is ignorant of this identity can entertain the possibility of its being carried out does not show that it is possible; it shows only that he does not know that it is not.

V. THE SOUL AS LIFE AND THE SOUL AS THOUGHT

It is useful in concluding, I think, to compare the philosophical conception of the mind with what was once the philosophical conception of life. It was once pretty much taken for granted that men and other animals *possess* something which inanimate things lack, namely, life, and that it is *because* they possess this that they can do all sorts of things that inanimate things cannot do, such as move themselves, assimilate nourishment, reproduce their kind, and so on. Aristotle classified the souls of living

things according to the abilities they imparted to their owners, and thought that even vegetables had souls. Indeed, an animal's *life* and *soul* were generally thought to be one and the same thing. The very word "animal" has its origin in this belief. Socrates, according to Plato, was even able to convince himself of his own immortality on the basis of this notion for, he thought, if it is only because he *has* a life or soul to begin with that he is a living man, then it is idle to fear the death of that very soul. Life seemed to him identical with his soul, but accidental to his body, indeed even foreign to such a thing of clay. A similar model was at work in Descartes' philosophy when he declared that the soul could never stop thinking. Thought seemed to him identical with his soul, but positively foreign to his body.

Now of course we still talk of life that way, but we no longer take such common modes of speech as descriptive of any reality. We speak of a man "losing" his life, of a man "taking" another's life, of the "gift" of life, and even of the "breath" of life which God is supposed to infuse into an otherwise *lifeless* body. But these are plainly metaphors. No one supposes that a man or animal moves, assimilates nourishment, reproduces, and so on *because* it is possessed of life. We no longer think of life as something added to an animal body, some separable thing that quickens matter. To distinguish something as a living animal is only to call attention to the very complicated way the matter of its body is organized and to a large class of capacities which result from such organization. A living body is simply one in which certain processes, some of them frightfully complex and ill understood, take place. A living body, in short, differs from a non-living one, not in what it possesses, but in what it does, and these are facts about it that can be verified in a straightforward way.

I have been urging a similar way of speaking of the mind; not as something mysteriously *embodied* here and there, and something that is supposed to *account* for the more or less intelligent behavior of certain beings. A being capable of more or less intelligent thought and action differs from one lacking such capacities, not in something it possesses,

but precisely in what it does. And this, incidentally, explains why a man tends to regard it as a deep insult to be told that he has no mind. It is not because he is thus divested in our eyes of some possession dearly prized, but rather, because such a remark is quite rightly taken to mean that he lacks certain important and distinctively human abilities and capacities. If a man is assured that his possession of certain more or less intellectual abilities is in no way in question, he feels divested of nothing upon learning that among his parts or possessions there is none that is properly denoted "a mind."

VI. DOES MATTER THINK?

Probably every philosopher has felt more or less acutely at one time or another a profound puzzlement in the idea of (mere) matter doing those various things rightly ascribable only to persons. How, it is wondered, can a body think, deliberate, imagine things, figure and plan, and so on?

This is really no proper source of bafflement, however. No one can say, *a priori*, what the highly organized material systems of one's body are or are not capable of. It was once thought incredible that matter, unquickened by any soul, could be alive, for matter seemed to inquirers to be inert or lifeless by its very nature. Yet we see around us all the time specimens of living matter—in the merest insects, for instance—so philosophical prejudice has had to yield to the fact. Similarly, I submit, we see around us all the time specimens of thinking matter; that is, material beings which deliberate, imagine, plan, and so on. For men do in fact do these things, and when we see a man, we are seeing a material being—a dreadfully complex and highly organized one, to be sure, but no less a visible and palpable object for that. In any case, the seeming mystery or incredibility that may attach to the idea of matter exercising intellectual capacities is hardly dissolved by postulating something *else* to exercise those capacities. If there is a difficulty in comprehending how a body can do such things, there is surely no less difficulty in seeing how something which is not a body can do them any better.

VI: RELIGIOUS VALUATIONS OF SCIENTIFIC TRUTHS

WILLIAM A. CHRISTIAN

MOST philosophical discussions of religion deal with the conceptual problems about traditional theism. But, even to handle some of these, comparable doctrines ought to be taken into account. "He who knows one, knows none"¹ has a point, for a certain sense of "know." We have to know more than one to know, in that sense, some particular one. Along with other things we can learn if we do not fix our minds on familiar doctrines only, we can get more light on the logic of religious doctrines including the familiar ones.²

I shall aim at this within the bounds of my topic, and I begin by putting its main point in a preliminary way. A scientific proposition has, or ought to have, religious import for someone if, from some religious doctrine he accepts, there follows a positive valuation of knowing its truth. Saying the valuation follows from the doctrine means that, if he accepts the doctrine, he would be acting inconsistently if he failed to make the valuation. So the cases where the proposition ought to have religious import, but does not, are cases where someone is being inconsistent.

Sometimes a religious doctrine has been thought to depend for its truth on some specific scientific theory, for example on Ptolemaic astronomy in the 17th century and on Aristotelian biology in the 19th century. I shall not discuss cases of this kind. I shall discuss only cases where valuations of scientific truths follow from doctrines which do not themselves depend for their truth on specific scientific theories.

A scientific truth can be valued in nonreligious ways also, as when knowing it helps to satisfy some moral obligation or when its simplicity and elegance give enjoyment. These valuations, like religious valuations of scientific truths, are not judgments that the scientific proposition is true. They are judgments that knowing its truth contributes to a moral satisfaction or to an aesthetic satisfaction.

Similarly a positive religious valuation of a scientific proposition is a judgment that knowing its truth contributes to a religious satisfaction. The activity of formulating a scientific proposition springs from a scientific interest, and a decision that the proposition is true or untrue is not a moral or aesthetic or religious judgment but a scientific judgment.

Historically religions have varied in their valuations of the natural sciences. A detailed comparative study of these would be very rewarding. It would also be a formidable undertaking because the doctrinal schemes of the major religions, Hinduism, Buddhism, Judaism, Christianity, and Islam, are very complex and they are still in process of development. Also other doctrines, for example those of the Stoics, Plotinus, Spinoza, and Comte, ought to be taken into account. I shall not even undertake a sample historical comparison, since I want to make some logical points, not historical ones. So I shall develop some hypothetical forms of religious doctrines and show how valuations of scientific truths would follow from some of them. I believe their relevance to various historical schemes of doctrines will be reasonably clear.

First let us construe the functions of scientific propositions and of religious doctrines in certain ways. Other ways of construing them may be possible, with other consequences.

I

Let us construe the natural sciences as inquiries aimed at answering the question: What, exactly, are the uniformities in the apparent world? Then observations, measurements, predictions, and empirical generalizations are important because, on one hand, they may suggest uniformities and, on the other hand, they may either support statements of uniformities or render them less plausible. The exactness aimed at calls for mathematical formula-

¹ Max Müller's aphorism about religions, adapted from a remark by Goethe about languages.

² Among recent studies in this vein are Ninian Smart, *Reasons and Faiths* (London, Routledge and Kegan Paul, 1958) and J. M. Bocheński, O.P., *The Logic of Religion* (New York, New York University Press, 1965).

tions of invariant relations, though a science may have to go through various stages before it is ready for them, though new mathematical concepts may have to be discovered, and though these formulations may hold only approximately. The apparent world is nature, in that sense of the term which Whitehead explains in *The Concept of Nature* as what is "there for sense-awareness" (p. 14). It is a domain of experience which expands as more powerful microscopes, telescopes, and other devices are invented. So far as the argument of this paper is concerned, I believe it does not matter whether we think of the apparent world in a phenomenalist way or in a realistic way.

Next let us construe religious doctrines as follows. In any religious community which becomes reflective there develops an attempt to explicate its beliefs in a more or less systematic manner. Further, for any religious community, there is some starting point for its thought. This may be an experience of some type (for example of satori, or of an access of moral power) or some historical event (for example the enlightenment of Gotama, or the life, death, and resurrection of Christ), to which the life of the community is a response. The development of doctrines proceeds from this starting point, which gives the enterprise its unity and continuity.

We might not want to call this theology in every case, because some religious communities are not theistic in any clear sense. We might want to restrict the term theology to mean the development of doctrines in some theistic religious community. I shall observe this restriction.

Thus the development of religious doctrines is rooted in religion, and a particular development is rooted in some particular religion. So we should ask whether there is some question, or some family of questions, to which various systems of religious doctrines might be thought to be addressed.

As we study various religious systems in their theory and practice it appears that they involve pointing to something or other which is valued in a certain way. They point to something taken as holy, or as the supreme goal of life, or as the chief end of man, or as the source of being, or as the way to freedom from inner bondage, or in some other such way. So religious doctrines can be construed as answers to one or another of a family of questions, and we can express the common intent of these questions as follows: What is it that is of central and ultimate significance for human life? So the form of

a basic religious question could be expressed by the open sentence "What is it that is P ?", using the letter P as a variable and allowing, as values of this variable, predicates such as "holy," "the supreme goal of life," "the source of being," and others. Questions of this form ask for a reference to something of which the predicate is true.

Since schemes of religious doctrines involve different referential terms, "God," "Allah," "Nature," "Brahman," "Nirvana," "Enlightenment," "Humanity," and others, as well as different basic predicates, I shall use the letter m as a variable to stand for that of which, in some scheme of doctrines, some basic predicate is understood to be true. So m stands for something or other which is valued in a distinctively religious way.

If we construe the natural sciences and religious doctrines in these ways, we are looking for connections between an enterprise which addresses itself to the question "What, exactly, are the uniformities in the apparent world?" and an enterprise which addresses itself to a question of the form "What is it that is P ?"

II

Now let us consider some possible doctrines (strictly speaking, forms of doctrines) about the relation of something m , something valued in a distinctively religious way, to the apparent world.

(A) m is a process in the apparent world.

The proponent of Doctrine (A) identifies m as³ a process in the apparent world, for example as the *nisus* toward deity (S. Alexander) or as the holistic process (J. C. Smuts) or as the process of progressive integration (H. N. Wieman). Then there might be a religious doctrine predicating that some uniformities or other, which could be formulated mathematically, hold true of m . Later I shall call such doctrines U doctrines.

Against this it could be objected that such a predicate would fail to convey the attitudes which are appropriate in religious discourse. Hence these could not be religious doctrines. This may not be a decisive objection. Taking an analogous case, we would not hesitate to react with genuine and profound emotion to a musical performance, though we would allow that some laws of physics apply to the performance. Again, it is both natural and not unreasonable to be morally responsible to other

³ One might identify someone as a lawyer, or as the president of the First National Bank, though there are many other things one would want to say about him also, by way of further identification or otherwise.

persons, though we would allow that many invariant relations, which can be formulated mathematically, hold true of their behavior. In our case, while the predicate taken by itself would fail to convey the subjective form of a religious attitude, the statement as a whole, involving as it does a reference to something valued in a distinctively religious way, might do so, and thus qualify as a religious doctrine.

(B) *m* is the total structure of the apparent world.

Doctrine (B) identifies *m* as the total structure of the apparent world. Then it would seem that *m*, or at least some aspect of *m*, would fall within the scope of science, and that formulations of invariant relations holding in the apparent world might have religious import. Before asking just how scientific propositions might be given religious import by way of (A) or (B), I shall introduce a third way of identifying *m*. Then I shall discuss these three cases together before going on to some others.

(C) *m* is an agent acting in the apparent world.

Doctrine (C) identifies *m* as an agent acting in the apparent world, whether locally and occasionally or pervasively and continually. Then these actions, or rather the events in which these actions are manifest, would seem to fall within the scope of science, at least in the sense that scientists would be committed to looking for uniformities in these events as in all others. Thus if there had been a physician in Bethany he would have been committed to looking for uniformities in the sequence of events which included the death and resurrection of Lazarus.

This brings up the subject of miracles and I shall deal briefly with the ordinary modern concept of a miracle. Putting historical questions aside, I am concerned only with conceptual questions, with criteria for saying of some event that it is a miracle.

According to the ordinary modern concept of a miracle, a statement of the form "*e* is a miracle" means in part⁴ that the event cannot be interpreted by any laws of nature whatever and is in this way inexplicable. "Whatever" removes restrictions to laws presently established at the time of the utterance, or to laws which might be formulated and established using concepts currently at hand. With either of these restrictions it would be easy, indeed too easy, to satisfy the criterion. When R. F. Holland⁵ speaks of miracles as events which are

"conceptually impossible" I am not sure whether he is removing these restrictions. Certainly there are circumstances when it would be "natural enough" (Holland's phrase) to say "I cannot conceive a framework within which *e* would fit."

This is not a Biblical concept, because laws of nature did not figure in the thought of Biblical times. It goes back at least as far as Spinoza, who thought it involved an absurdity. Hume's attitude to this concept in his essay on miracles is not very clear. Other concepts of miracles have been defended by various theologians, for example by Schleiermacher.

This is a very strong criterion indeed. To appreciate its strength, notice that a negation of "*e* is inexplicable" would not entail that the event was within human power to perform. It might be both explicable, in the sense that some formulations of uniformities hold true of it, and beyond human power. Nor would the negation deny that *e* is a unique event, unduplicated in the course of human events or indeed in cosmic history. Nor would it deny that in *e* a *mysterium tremendum et fascinans* can be divined.

It seems to me that statements of the form "*e* is inexplicable," when they bear the sense I have mentioned, and with them the ordinary modern concept of a miracle, are not just improbable but absurd. They may or may not be logically absurd, as Spinoza thought they are, though there are plenty of logical difficulties about them. But they do seem to be absurd in the following way.

Suppose that all human beings are potentially capable of engaging in scientific enterprises and that they have responsibilities, varying with their powers and circumstances, for advancing these enterprises in some way. Then people who are not scientists by vocation would have a personal stake in what scientists are doing, because their interests would commit them to these enterprises, at least by way of sympathetic participation. (We might suppose, similarly, that all human beings have potentialities and responsibilities for advancing religious understanding for the sake of practical wisdom.)

Now if it is impossible to find any uniformities which would hold true of some event, this implies it is absurd to look for any. But from the standpoint of a scientific interest this implication is itself absurd because a scientific interest commits one to looking for uniformities throughout the apparent world.

⁴ Additional criteria, for example that the event is very surprising, or that it manifests divine power, need not be discussed here.

⁵ "The Miraculous," *American Philosophical Quarterly*, vol. 2 (1965), pp. 43-51.

The proponent of Doctrine (C) might view this situation with equanimity. He might say that interpreting an event as an action of m does not depend on a failure (for whatever reason) to find uniformities which apply to it. And he might say further that true scientific propositions about the event would contribute to an understanding both of the mode and the setting of m 's activity in the apparent world.

III

Some general considerations are in order before returning to Doctrines (A), (B), and (C) in particular. We are asking how scientific propositions might take on religious import. Certainly this does not mean they would turn into religious propositions. The principles in accord with which we judge them true or untrue are independent of any religious import they might acquire, and these principles differ from principles for judging religious doctrines.⁶

Now consider the following proposition, which I shall refer to as S :

Some uniformities or other, which can be formulated mathematically, hold true in the apparent world.

This is not a scientific theory but instead a supposition involved in having a scientific interest, so it is basic to any scientific enterprise. How might this basic supposition of science acquire religious import? This could happen if a doctrinal scheme required its adherents to value knowledge of the apparent world in a positive way, and a valuation of this sort would follow from a doctrine of any of the following forms and perhaps others:

Knowledge of the apparent world can contribute to understanding of m (or, to realization of m ,⁷ or effective obedience to m). I shall call such doctrines K doctrines. If anyone accepts a K doctrine he is thereby committed to a positive valuation of knowledge of the apparent world.

Thus for example, acting on a K doctrine whether formulated or not, a priest might study the stars and the phases of the moon to learn the propitious time, the time favored by the gods, for a sacrifice or a military expedition.⁸ A teacher of yoga might take

account of various bodily reactions to prescribe postures favorable to spiritual concentration. And a theologian might argue as follows: (a) Since God is the creator of the world around us, we ought to know it better so as to understand his creation, and (b) God commands us to love our neighbors as well as ourselves and, since this involves action in the apparent world, we need to know this world to do so effectively. The religious value of knowledge of the apparent world might be for understanding or practice or both, as these examples suggest.

Acceptance of a K doctrine would not necessarily involve having a scientific interest and thus actually making the basic supposition of science. So S itself does not follow from a K doctrine. What does follow is that we ought to have knowledge of the apparent world. But this consequence could be satisfied by what we learn in ordinary experience, without involving S .⁹ No religion, I suppose, would want to exclude people living in a pre-scientific culture though it might want to educate them.

At the same time, since a K doctrine would lend religious import to any inquiry about the apparent world, and thus to the basic supposition of such an inquiry, it would lend religious import to S , if S is asserted. So the connection between a K doctrine and S is not that S follows from a K doctrine. It is instead that a positive valuation of what is supposed in S follows from a K doctrine, if S is asserted. If some proponent of a K doctrine is led by experience to suppose S then, since an inquiry guided by S aims at knowledge of the apparent world, he is bound by the K doctrine to make a positive religious valuation of what is supposed in S , and of the results of the inquiry guided by it. If he fails to do so, he is being inconsistent.

Now consider some consequences of Doctrines (A), (B), and (C) in particular. As we have noticed, in the presence of Doctrine (A), it could be an acceptable doctrine that some uniformities, which can be stated mathematically, hold true of the process m . (I say it could be, because I am speaking of hypothetical doctrinal schemes.) In the presence of Doctrine (B), it could be an acceptable doctrine that some uniformities, which can be stated mathematically, hold true of the structure of the apparent world. In the presence of Doctrine (C), it

⁶ I have discussed a number of principles of religious judgment in *Meaning and Truth in Religion* (Princeton, Princeton University Press, 1964).

⁷ If m is an experienceable quality or state as in Doctrine 'E', which we shall come to later.

⁸ See Gerardus van der Leeuw, *Religion in Essence and Manifestation*, tr. by J. E. Turner (London, 1938), ch. 54 on divination, "which at one time has a calm and almost scientific character and at another a more ecstatic form," p. 379.

⁹ Unless we define knowledge in an unduly restrictive way.

could be an acceptable doctrine that some uniformities, which can be stated mathematically, hold true of the events in which the agent m acts.

Let us refer to such derivative doctrines as U doctrines. Some U doctrine might be only on the periphery of the scheme it belongs to, or it might be more central to the scheme, as for example in a Spinozistic version of Doctrine (B). Its links with the main doctrines in the scheme might be loose and weak, or they might be firm and strong. Since all these derivative doctrines involve the basic supposition of science—though it does not entail any of them, S would acquire a further and more specific religious import in each case.

No specific scientific conclusion could be deduced from any of these U doctrines, nor of course from the basic supposition of science. But specific scientific truths, as well as the basic supposition of science, would acquire religious import. The reason is that from a U doctrine some K doctrine or other follows. That is to say, if m is identified as in (A) or (B) or (C), and if some uniformities hold true of m or m 's actions, then it follows that knowledge of the apparent world (specifically of these uniformities in the apparent world, whatever they are) can contribute to an appropriate response to m .

So if a doctrinal scheme requires a U doctrine, someone who accepts that scheme is thereby committed to some K doctrine also. And that means he is committed to valuing, in a positive way, knowledge of the uniformities which hold true of m (or ...). We might go further and say that, since the concepts employed in formulations of uniformities belong to a network of scientific theories, he is committed to approving the scientific enterprise as a whole.

The proponent of Doctrine (A) might explain this as follows: "We acknowledge a certain process as the source of our life; we conform our actions to its operation; and we trust it to complete the intent of our actions, so far as they can be completed. A specific understanding of the regularities which hold true of this process would guide us to a more concrete, and hence a more practical, understanding of that which we value, conform to, and trust."

The proponent of Doctrine (B) might say: "We think of ourselves as parts of the total structure of the apparent world. We accept our limitations and disappointments as necessary consequences of that structure, not with a grudging resignation but willingly and freely. Knowing how laws of nature hold true would give us a more concrete understanding of that structure, and of the necessities of

which our limitations and disappointments, as well as our powers and achievements, are the consequences."

The proponent of Doctrine (C) might say: "We view events in the apparent world, or some of them at least, as expressions of m 's active power, as acts of justice and mercy. Knowing specific invariant relations which hold true of these events would give us a more concrete understanding of the situations in which m acts, and how m acts in them."

In this discussion of U doctrines I am not giving theological or other arguments for them. Nor is it essential to show that they have been asserted, though I have mentioned some interesting historical developments. My object is to call attention to their possibility and to show how they are possible.

In (A) and in (B) and in (C), m is identified in ways which give footholds for U doctrines. In each case a doctrinal scheme could permit some U doctrine. But other doctrinal schemes such as those introduced by Doctrines (D) and (F) do not permit U doctrines. These identifications of m do not seem to afford any footholds for such doctrines. Yet as we shall see K doctrines are possible in (F) though not in (D).

So, though from any U doctrine some K doctrine follows, the converse does not hold. It would be consistent to say that knowledge of the apparent world can contribute to an appropriate response to m , asserting a K doctrine, and at the same time to say that no uniformities which can be formulated mathematically hold true of m , denying U doctrines. In this case it is still possible for scientific truths to take on religious import by way of valuations following from the K doctrine. U doctrines are sufficient but not necessary conditions for positive religious valuations of scientific truths.

IV

Other doctrines about something m , something valued in a distinctively religious way, and the apparent world, are theoretically possible.

(D) m is unconnected with the apparent world.

The proponent of Doctrine (D) places the concept of m in an entirely different domain of discourse from discourse about the apparent world, in which scientific concepts find their place. Discourse about m and scientific discourse have no positive connections. There would be no foothold for a U doctrine, and K doctrines would be denied.

He might say in explanation (though other ways

of developing (D) are possible) that discourse about *m* expresses (which is not to say it reports or describes) certain attitudes or emotional states or moods. This would be true, so far as it goes, of religious doctrines generally. But the proponent of (D) would mean much more; the expressive significance of religious doctrines (at any rate of *his* religious doctrines) is their only significance.

Then scientific truths could have at most only a negative import for religion—negative not by way of logical conflict but possibly in two other ways. They might have a negative significance by way of contrast; they show what religious doctrines are not. Also, apprehensions of impersonal truths about the world might be thought to conduce to some mood, for example despair, which is thought to be a negative condition of faith. Or, instead of having a negative import for religion, science might have no import at all.

(E) *m* is a state attainable in the apparent world.

Though the history of religions has been suggestive I do not claim there are historical examples of this and the other doctrines I have been constructing, nor do I need to. My object is not to generalize from historical forms of religion but to construct hypothetical cases which can help bring out the conditions under which valuations of scientific truths are derivable.

This doctrine might be developed in various ways including the following one. We tend to clothe the objects of perceptual experience in myths and symbols woven from our own inner conflicts and frustrations. Thus we fail to apprehend things as they are and for what they are. We need to liberate ourselves from these distortions of experience.

We are liberated from them, at least for the time being, whenever we encounter a particular thing or situation as just itself. The Zen saying, "The blossom is red, the willow is green," is meant to suggest the none-other-ness of the blossom and the willow, though of course much more also. When we experience something as just what it is our unregenerate cravings and hostilities are brought up short and our inner derangements tend to be dissolved.¹⁰

The state of liberation might also be described as being awakened, since it is something like awakening from a dream, or as enlightenment. It involves among other things getting rid of various kinds of illusions in our experience of the apparent

world, which are both products and causes of our own unhappiness. In any particular case the liberation is more or less important corresponding to the importance of the illusion in emotional life.

It is then fairly clear how a *K* doctrine would be in order and how specific scientific propositions could have import for religious experience. Inasfar as we know anything about something we are liberated from some illusion about it. Whether Doctrine (E) permits a *U* doctrine would depend on further developments. Certainly it permits saying that some uniformities, which can be stated mathematically, hold true of the process by which *m* is attained even if none apply to *m*.

(F) *m* is the Form of the forms in the apparent world.

In Doctrine (F), *m* is identified as a reality beyond the apparent world but—unlike Doctrine (D)—as beyond the apparent world in a determinate way. It is that formless Form expressed and adumbrated by all forms, by some more truly than by others, more by scientific truths than by perceptual truths, and more by logical truths than by scientific truths.

The basic supposition of science would gain religious import by way of a *K* doctrine together with other doctrines giving the apparent world its place in the hierarchy of being. Also specific scientific truths would gain religious import naturally and directly, since an apprehension of some specific uniformity holding in the apparent world would lure the mind toward the Form beyond forms. Thus as with (A) or (B) or (C) or (E) scientific knowledge could enhance religious experience, and religious experience could contribute incentives for scientific activity. It seems that no *U* doctrine could be permitted.

V

We have been exploring ways in which positive valuations of scientific truths are derivable from religious doctrines. Now we can schematize these derivations in the following manner:

(1) *m* is *P*.

Here, one function of a predicate substitutable for *P* ("holy," "the supreme goal of life," "the ground of being," and others) is to assign to *m* some primacy or other in the concerns of life.

(2) The appropriate responses to *m* are *r*₁, *r*₂, *r*₃...

What constitutes an appropriate response to *m* will be determined by the interpretation of *P* and by

¹⁰ In such a context Bishop Butler's saying, "Everything is what it is, and not another thing," would acquire an unexpected religious significance.

various other doctrines. A doctrinal scheme may thus enjoin understanding of m , realization of m , obedience to m , or other responses. From a doctrine of form (1), some doctrine or other of form (2) must follow. Otherwise m would not have an effective primacy in the concerns of life.

(3) m has relation R to the apparent world.

Doctrines (A)–(F) are possible interpretations of R , though not the only ones. In the case of a particular religion such a doctrine would refer, for its justification as a doctrine of that religion, to the starting point of the doctrinal scheme.

(4) Knowledge of the apparent world can contribute to an appropriate response to m .

I have called doctrines of this form K doctrines and suggested how they follow from certain developments of Doctrines (A), (B), (C), (E), and (F), though excluded by Doctrine (D). (If the other doctrines were developed differently K doctrines might not follow from them either.) It would follow that knowledge of the apparent world is a good thing to have.

(5) The natural sciences produce truths about the apparent world, namely exact statements of uniformities in it.

(6) Scientific truths can contribute to an appropriate response to m .

This follows from (4), where (4) is valid, together with (5). From this it would follow that scientific knowledge is a good thing to have.

Now (6) would be strengthened if:

(7) Some uniformities, which can be stated mathematically, hold true of m (or, of m 's

actions in the apparent world, or, of the process of attaining m).

I have called doctrines of this form U doctrines and suggested how they might be congruent with Doctrines (A), (B), (C), and possibly (E). Though neither (4) nor (6) requires (7), if some doctrinal scheme permits (7) then scientific truths could contribute in a more direct manner to an appropriate response to m than if (7) were not permitted.

In conclusion, if we ask whether, and how, science can have religious import, the right answer will depend on the scheme of religious doctrines we have in mind. I have introduced only a few possible schemes, by way of doctrines relating something m to the apparent world. These are the key doctrines for the problem at hand. Each of these schemes could be developed much further and other schemes could be developed. One reason why the doctrinal schemes of the major religions are more complex is that theologians (and their equivalents) often find themselves identifying something m in more than one way, as they develop doctrines from some starting point, so different strands of thought must be woven together.

None of this implies that any religious doctrine depends for its truth on any scientific theory. Some philosophers have argued that the basic supposition of science and therewith the enterprise of natural science as a whole depends on, or in some other way requires, some religious doctrine. This seems doubtful to me but in any case my argument is independent of that issue. Also, I have not considered the relation of speculative philosophy to religious doctrines and scientific truths.

VII: UNIVERSALS AS DESIGNATA OF PREDICATES

ROLF EBERLE

IN discussions concerning the traditional problem of universals, one frequently encounters the position that predicates designate in much the same manner in which names name, except that the typical designata of predicates (called "universals") differ, in kind, from the typical designata of names (called "individuals"). We shall attempt a clarification and a brief discussion of this position. The clarification will take the form of principles which may govern extensions of the naming relation from names to predicates. The notions of a universal and of an individual will then be characterized, and various positions of realism and nominalism will be distinguished, with respect to the semantical relations which can be employed in interpreting predicates. Thus, realism and nominalism are here treated as proposals for constructing alternative theories of reference. A brief and informal discussion of some of these theories will be given.

Consider the sentence "Socrates is human." The name "Socrates," we shall suppose, designates the person Socrates. But what, if anything, might be designated by the predicate "human"? Different answers to this question distinguish realists and nominalists who raise the problem of universals in the context of a theory of reference.

Realists, with respect to this issue, have tended to maintain that each meaningful predicate, like every proper name, designates some unique entity. The relation obtaining between a predicate and its designatum has been conceived in close analogy to the naming relation, except that the designata of names and the designata of predicates have been placed into different categories.¹ Among suitable designata of predicates, realists have listed properties and classes. Thus, the predicate "human" would be taken to designate either the property of being human or the class of all humans.

On the other hand, nominalists who have

approached the problem of universals in this manner have tended to favor instead one of the following alternatives: (i) The position that predicates fail to designate altogether. On this view, predicates are syncategorematic expressions: Instead of entering a semantical relation akin to the naming relation, predicates receive only an indirect interpretation through truth conditions specified for all sentences in which these predicates may occur. This manner of construing predicates has been vigorously defended by Quine.² (ii) The position that predicates indeed "designate," but that this semantical relation is structurally different from that obtaining between a name and the thing named. For example, predicates may be construed as having multiple denotation. Thus, the predicate "human," on this view, is taken to "designate" each particular human being. Among contemporary writers who have made extensive use of such relations of designation, we mention especially R. M. Martin.³ (iii) The position that predicates designate in exactly the same sense in which names name, even in the sense that the designata of predicates are the same sort of entities also named by proper names. On this view, and in a sense yet to be clarified, predicates designate individuals.

In the statement of each of these positions, a notion of designation is needed which may apply to predicates, and which is sufficiently clear, in the relevant respects, to afford a comparison with the naming relation. This latter relation, in its applications to simple and compound names, has received careful investigation.⁴ For this reason, we shall take for granted an understanding of the naming relation in so far as it applies to names (and proper descriptions). But the present discussion demands a notion of designation which is appropriately extended from names to predicates, and which is so specified that individuals and universals can be distinguished just

¹ This interpretation of realism appears, for example, in Bertrand Russell, "Problem of Universals," *Polemics*, vol. 2 (1946), pp. 21-35.

² W. V. Quine, "Designation and Existence," *The Journal of Philosophy*, vol. 36 (1939), pp. 701-909.

³ R. M. Martin, *Truth and Denotation* (Chicago, 1958).

⁴ Rudolf Carnap, *Meaning and Necessity* (Chicago, 1947), especially ch. III.

in terms of certain features of the designation relation. We make it our first task to characterize such a notion of designation.

To begin with, any extension of a naming relation from names to predicates must satisfy the obvious requirement that it is a relation whose domain is a set of names and predicates and such that the result of restricting the extension to the set of names in its domain is the original naming relation.

Throughout this discussion we take for granted that the notions of a name and of a predicate are syntactical and not semantical notions. In particular, we shall not regard it as a contradiction in terms to speak of a denotationless name, or of a name which fails to name, since expressions can belong to the syntactical category of names without entering a semantical relation of naming. For this reason, we do not require that a naming relation interprets the entire set of names of a language, or that an extension of that relation interprets all names and predicates of that language.

It is usually regarded as an essential feature of a naming relation that no name bears that relation to more than one object.⁵ It can plausibly be maintained that any semantical relation which has claim to being akin to a naming relation must preserve this feature of being single-valued. Accordingly, the following condition may be imposed:

- (C₁) Any proper extension of a naming relation from names to predicates is a (single-valued) function.

A relation which conforms to C₁ will be called "single-valued designation"; one which fails to meet C₁ will be called "multiple-valued designation." In saying that multiple-valued designation is a relation which lacks one of the essential features of a naming relation, we do not imply that this relation is unsuitable for the purpose of interpreting predicates.

According to the second nominalistic position, mentioned above, each predicate "designates" all objects of which it is truly predicable. From this point of view then, predicates enter typically a relation of multiple-valued designation, a relation which is in this respect unlike the naming relation.

It is a further characteristic of a naming relation interpreting the names of a given vocabulary that every object named is of such a nature that it could always also be named by any other name in the same vocabulary. Thus, if N_1 and N_2 are names naming, according to a given naming relation, the

respective objects x_1 and x_2 , then that relation which coincides with the given one, except for assigning x_2 to N_1 and x_1 to N_2 , is also a naming relation. Generalizing from this example, let us say:

- (1) A class K of binary relations is closed under interchange of second relata just in case the following condition is satisfied: if R is any relation in K relating X with x and Y with y , then that relation R' which differs from R at most by relating X with y and Y with x is again in K .

Thus, the class of naming relations is closed under interchange of second relata. This feature of naming relations is well known but not made explicit in treatments of those relations. If extended relations of designation are to be similar to naming relations in this respect, then one should be able to interchange any two objects assigned to names, any two objects assigned to predicates and, more significantly, any two entities one of which is assigned to a name and one of which is assigned to a predicate. More precisely expressed, the requirement that the class of extended naming relations should be closed under interchange of second relata has the following consequence:

- (2) If N is a name and P is a predicate, both belonging to the same vocabulary, and if N designates n and P designates p , according to the proper extension of a naming relation, then that relation which coincides with the given one, except for assigning p to N and n to P , should itself be a proper extension of some naming relation.

For example, if John is named by a name and if predicates designate in the same manner as names, then one should also be prepared to construe some predicate as designating John. Yet, a concrete object like John, one feels, is not the sort of thing which can properly be designated by a predicate, if predicates are to play their intended semantical role. Intuitively, there is a difference between entities which are proper designata of names and ones which are proper designata of predicates. As Strawson⁶ and others have noted, this difference is not one which can be expressed in purely syntactical terms. If it is felt that predicates cannot designate concrete things, this feeling is not prompted by consideration of the notational characteristics of predicates or of their position in sentences. Instead, the intuitive difference between names and predicates should be explained in semantical terms; and

⁵ Carnap calls this "the principle of univocality," *ibid.*, p. 98.

⁶ P. F. Strawson, *Individuals* (Garden City, 1959), ch. V.

to this purpose we avail ourselves of the customary notion of a universe of discourse.

Assuming that a given language contains names, a universe of discourse comprises just those entities which are assigned to the names of the language by some naming relation into that universe. In this sense, a universe of discourse can be regarded as a set of entities which are construed as "nameable" for the purposes of a given discourse. In as much as any set whatsoever can be a universe of discourse, no restriction is imposed on the nature or number of entities which may be construed as nameable for the purposes of any given discourse. It might appear to follow that every entity which is fit to be ascribed to some predicate of a language is also fit to be ascribed to a name of that language. However, if predicates are taken to designate classes or properties of nameable things, then not every entity which is fit to be ascribed to a predicate can itself be nameable, for the following obvious reasons:

By an application of Cantor's Theorem to an arbitrary universe of discourse, there are fewer elements than sets of elements of that universe. Thus, no matter which set is taken to comprise all nameable things, there are always more sets of nameable things than nameable things. Furthermore, the notion of a property is usually characterized in such a fashion that entities in distinct sets have distinct properties, whereas distinct properties may be common to all objects in the same class or extension. Given such a principle of individuation for properties, there will be even more properties of nameable things than classes of nameable things, for any universe of nameable things. If, as some realists would have it, all classes or properties of nameable things are fit to be ascribed to predicates, then some entities which qualify as designata of predicates cannot qualify as designata of names.

Furthermore, some universes of discourse lack this feature: that every element of the universe is itself a set of elements of that same universe (that is, some sets are not transitive or complete). If predicates are taken to designate sets of elements of some such universe, then not every object named in that universe is also a class fit to be ascribed to some predicate. Thus, if predicates designate sets of nameable things, then it is not generally possible either to construe designata of predicates as nameable or nameable things as fit designata of predicates. Hence, there are formal, as well as intuitive reasons for the claim that the proper designata of names differ, in kind, from what realists regard as the proper designata of predicates.

It is the intuitive content of the second condition about to be stated that predicates cannot properly be said to designate in the manner of names, unless they designate the same sort of entities which one is also prepared to name by names:

- (C2) Every proper extension, from names to predicates, of a naming relation into a universe of discourse is again a relation into that same universe of discourse.

Given our understanding of what a universe of discourse is to be and assuming that a language comprises both names and predicates, the class of extended naming relations which satisfy C2 is just the class of extended naming relations which is closed under interchange of second relata. Hence, every relation of designation which satisfies C2 admits of an interchange of the designata of names and predicates expressed earlier under (2). The condition C2 and our earlier condition C1 will serve us in classifying proposed extensions of naming relations and will permit a more precise formulation of the various nominalistic and realistic positions mentioned previously.

According to realists, predicates enter a semantical relation which satisfies the condition C1, but not the condition C2, of an extended naming relation. That is, predicates have single-valued designation, but they do not, in general, designate the same sort of entities which are named by names. If predicates designate in the sense of C1 but not in the sense of C2, then their designata, whatever they be, may be called *universals*. By contrast, the elements in a given universe of discourse, the objects which are construed as being nameable, may be called *individuals*. According to nominalists, either of the following three alternatives are acceptable: (i) predicates fail to enter any extended naming relation and are, in this sense, syncategorematic; or (ii) predicates enter a semantical relation which meets the condition C2, but fails to meet the condition C1. In this sense, predicates are construed as having multiple designation, and as designating the same sort of entities which are also assigned to names; or, (iii) predicates are interpreted by extended naming relations which meet both of the conditions C1 and C2. In this sense, each predicate is construed as naming at most one nameable object. Thus, all three nominalistic positions would have it that either predicates fail to designate altogether, or that they designate individuals.

Nominalism can be understood as a demand for economy with respect to the categories of designata

of a language: either predicates should have no designata at all, or else they ought to designate the same kinds of entities which one is also prepared to name by names. By the semantical positions of nominalism here at issue it is left undetermined what sorts of entities one may choose to regard as nameable; but once this choice is made, one is urged not to draw from yet another category of entities, such as that of sets of nameable things or properties of nameable things, to supply designata of predicates.

In discussions concerning universals, the question is sometimes misleadingly raised whether predicates (really) designate, or whether they (really) designate entities of this or that sort. But there is surely nothing in the nature of predicates themselves which would make them either fit or unfit for designation. Nor can one settle the question whether there are universals, as we conceive them here, by searching throughout the universe for entities which might bear the unmistakable marks of being designated by predicates. The problem is rather this: whether an overall adequate semantics for various theories can be given (and notably, whether reasonable truth conditions can be specified for all sentences of subject-predicate form), if predicates are construed in the manners suggested by realists or nominalists. This semantical issue between realists and nominalists, we submit, is appropriately discussed by constructing alternative theories of reference and comparing their relative merits. Economy with respect to the kinds of entities designated by expressions may be one of these merits. Accordingly, we make it our next task to inquire whether there are advantages in construing predicates as designating expressions. Only one-place predicates will be considered.

Consider how the predicate "true" can be defined for atomic sentences of subject-predicate form. If the number of such sentences in a given language is finite and if there exist translations of all names and predicates in the meta-language, then a definition of truth can be given by conjoining all instances of Tarski's material criterion of adequacy.⁷ However, if a language should comprise infinitely many names or predicates which may apply to distinct entities, then it is no longer possible to define the predicate

"true" by listing, in a definition of finite length, all atomic instances of that criterion. Instead, one characterizes truth in terms of an auxiliary notion of designation which is defined on the infinite categories of names or predicates. For German sentences of subject-predicate form such a definition of truth may read:

- (3) If N is a name and P is a predicate, both of German, then the concatenate of N , "is," and P is a true sentence of German if and only if the designatum of N is a member of the designatum of P .

The intuitive adequacy of (3) presupposes that predicates are construed as designating their extensions.

Supporters of the position that predicates fail to designate cannot avail themselves of truth definitions such as (3). They face difficulties, therefore, in defining truth for languages with infinitely many predicates. Furthermore, formulations like (3) do not require that the English meta-language contains translations of all German predicates. Thus, the number of primitive expressions needed in defining truth can be reduced if predicates are treated as designating expressions.⁸

Our discussion concerning the syncategorematic treatment of predicates serves to illustrate that proposals regarding the interpretation of predicates cannot be evaluated in the absence of proposals concerning the treatment of truth. We shall now introduce considerations which suggest that the very notion of designation at issue here can be a clear notion only in the framework of a theory of truth.

Consider again the truth condition (3) in which both names and predicates of German are construed as designating expressions. The predicates, in this condition, are interpreted by a relation of single-valued designation whereby each predicate designates its extension. Let us now introduce another relation, called "application" in contrast to "designation," as follows:

- (4) If P is a predicate of German, then P applies to x if and only if there is a name N of German such that N designates x and the concatenate of N , "is," and P is a true sentence of German.

Thus, predicates enter the relation of application as well as that of designation. However, whereas the

⁷ Alfred Tarski, "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica*, vol. 1 (1935), pp. 261-405.

⁸ Predicates must also be treated as designating expressions if one is to provide a semantical notion of validity or logical truth (in contrast to the syntactical notions of derivability or of being a theorem). For, a sentence is logically true only if it is true under every permissible assignment of entities to predicates. For the special case where the entities assigned to predicates are their extensions, this point is discussed by Quine. See W. V. Quine, *From A Logical Point of View* (Cambridge, 1940), pp. 115-116.

relation of designation employed in (3) is single-valued and has extensions in its range, the relation of application defined in (4) is typically multiple-valued and has members of these extensions in its range. If both of these relations are used in a given theory of reference, it may not be clear whether that theory conforms to realism (by virtue of its relation of designation) or to nominalism (because of its relation of application). The relevant difference between the two relations is this: The relation of designation is here employed in defining truth for sentences of subject-predicate form, whereas the notion of truth (and hence indirectly the notion of designation) is used in defining application. In this sense, the relation of designation is more fundamental than, or presupposed by, the relation of application. Thus, if predicates should enter more than one semantical relation, in a given theory of reference, then we cannot decide which of these relations is the one we have called "designation" unless we inspect the manner in which the notion of truth (or related notions, such as that of satisfaction) is introduced in that theory. Accordingly, one can decide whether a given theory of reference is realistic or nominalistic only if that theory provides a sufficiently explicit treatment of truth to clarify the systematic status of any referential relations into which predicates might enter. The truth conditions exemplified by (3) specify that

- (5) a sentence of subject-predicate form is true if and only if the designatum of the subject is a member of the designatum of the predicate.

Use of the relation of membership in this condition implies that a sentence of subject-predicate form is true only if the predicate designates some non-empty class. But it does not appear necessary to regard classes as the only possible designata of predicates. To achieve greater generality with respect to the kinds of permissible designata of predicates, it seems natural to replace reference to membership in (5) by reference to an arbitrary relation R satisfying the condition that

- (6) a sentence of subject-predicate form is true if and only if the designatum of the subject bears R to the designatum of the predicate.

We call such a relation R a *relation of predication*, and draw attention to the fact that R relates, not expressions, but entities designated by expressions.⁹

In discussions concerning universals, it is fruitful to observe that the choice of a particular relation of predication is correlative with the decision to construe predicates as designating entities of a certain sort. Thus, the selected relation of predication will be that of membership just in case predicates are construed as designating classes; and it will be that of "being characterized by" just in case predicates are treated as designating properties. Perhaps some nominalists prefer to regard predicates as designating concrete exemplars. Thus, the predicate "is one meter long" might designate the standard meter, each color predicate might designate a patch of standard color, and so forth. A relation of predication suitable to this approach should obtain, for example, between an object and the exemplar of the color white just in case that object is white. It might appear that some relation of matching or of resemblance could satisfy this condition. Conversely, if some relation of resemblance is employed in formulating truth conditions for sentences of subject-predicate form, then the designata of predicates must be thought of as the kind of entities which can enter such resemblance relations. Generally, given any particular relation of predication, the designata of predicates will be entities of the sort to which individuals can bear that relation.

Many of the discussions traditionally given of realism and nominalism can be reconstructed as discussions concerning the merits of various relations of predication for the purposes of semantics. In partial support of this contention, we shall consider one nominalistic and one realistic position, and we shall attempt a sketchy reconstruction of some typical discussions given of those positions.

One view, which is frequently regarded as characteristic of nominalism, is this: that various objects, to which a predicate properly applies, resemble one another without having anything "in common."¹⁰ From this point of view, predicates apply to concrete objects by virtue of certain resemblance relations between these objects, and not due to any relation which these concrete objects might bear to some abstract entity.

⁹ P. F. Strawson, *ibid.*, Pt. V, sect. 8. Strawson calls some such relations "non-relational ties." Strawson seems to insist, whereas we do not, that the terms in such ties be of different logical type. Relations of predication cannot generally be expressed in the language with respect to which truth is defined. Our relations of predication are not "relations" if that word is reserved for designata of relation expressions of the object-language.

¹⁰ For a detailed discussion of the resemblance theory of classification and of its criticisms, see I. M. Bochenski in I. M. Bochenski, A. Church, and N. Goodman, *The Problems of Universals* (Notre Dame, 1956), pp. 33-57.

Apparently, the criticism most frequently advanced against the position just mentioned is this: Since all concrete objects resemble one another in some respect or other, the relation of mere resemblance cannot serve for the purposes of classification. The required notion is not that of resemblance, but rather that of resemblance in a given respect (resemblance with respect to color, with respect to shape, and so forth). And in explicating what is meant by the expression "resembling in the respect *R*," reference must be made to a quality or class, namely the respect *R*, in regard to which the particular resemblance relation obtains. Critics take this to show that the given nominalistic position cannot avoid reference to universals and is therefore untenable.

We believe that the following two demands constitute the semantical counterpart of the nominalistic position mentioned above: (i) that predicates should enter a relation of designation which satisfies both of the conditions C1 and C2 (that is, each predicate should designate a unique individual), and (ii) that the relation of predication should be that of resemblance. For sentences of the form "*N* is white" a corresponding truth condition might read:

- (7) "*N* is white" is true if and only if the designatum of *N* resembles the designatum of "white."

The unique designatum of the predicate "white" must be an individual belonging to the same universe of discourse to which all white objects belong, and it must resemble all of these objects. An entity which satisfies these conditions, one might think, is some concrete object which plays with respect to the color "white" the same normative role which is played by the standard meter with respect to the length "one meter."¹¹

Let us attempt next to formulate a semantical counterpart of the criticism outlined above. Such a critic would point out that some predicate other than "white" may truly apply to the designatum of "white." Suppose, for instance, that the exemplar of the color "white" should be both white and square, and that all square objects resemble one another. Then, according to condition (7), all sentences of the form "*N* is white," where *N* designates

some object which is square but not white, would turn out to be true. In response to this difficulty, a nominalist might want to replace the relation of resemblance by the relation of resemblance with respect to color. But this latter relation serves only for sentences containing color predicates. In general, several relations are needed: resembling with respect to color, with respect to shape, and so forth. Heeding these suggestions, a nominalist might specify truth conditions of the following pattern:

- (8) if *P* is a color predicate, then sentences of the form "*N* is *P*" are true if and only if the designatum of *N* resembles, in color, the designatum of *P*, if *P* is a shape predicate, then sentences of the form "*N* is *P*" are true if and only if the designatum of *N* resembles, in shape, the designatum of *P*, and so forth.

However, the distinction between color predicates and shape predicates is presumably not a purely syntactical one. Hence, a syntactical description of a sentence can no longer determine which truth condition is applicable to that sentence. Thus, the conditions in (8) do not seem to be serviceable as clauses in a definition of the predicate "true."

These sketchy remarks may suffice to indicate that the nominalistic "resemblance theory," as well as its criticisms, gain in clarity and interest if that theory is interpreted as a proposal for employing certain relations of designation and of predication in semantics. Let us turn next to a cursory examination of one realistic position and one of its criticisms.

Realists have traditionally maintained that whenever objects are correctly classified in a given respect, there subsists, in addition to these objects, some feature or property, and the classification is warranted due to a certain relation which all classified objects bear to that property. A semantical counterpart of this position would comprise the two demands (i) that predicates should enter a relation of designation which satisfies the condition C1 but not the condition C2 (that is, each predicate should designate a unique entity which need not be an individual), and (ii) that the relation of predication should be that between a thing and its properties.

Apart from the objection that the relation between a thing and its properties is not clearly

¹¹ Even if all and only those objects which are white should resemble one another, a link is still needed between the predicate "white" and these objects, if their resemblance is to serve the purposes of predication. If this link is that of multiple-valued designation, then the appropriate relation of predication is not resemblance but identity. If the predicate "white" singularly designates the class of these resembling objects, then the position in question is a realistic one. Hence, the demand that the relation of predication shall be that of resemblance can be combined with nominalism only if the predicate "white" designates a unique exemplar taken from this class of resembling things.

understood, the chief traditional criticism of the given position seems to be just this: that according to this theory predicates make reference to entities which do not really exist. Our previous remarks have paved the path to a semantical interpretation of this objection. Sentences of the form "There exists a so-and-so" are taken to be true just in case some object in the universe of discourse is so-and-so. And if *N* is a name, then sentences of the form "*N* exists" are plausibly regarded as true just in case *N* designates some object in that universe.¹² If realists ascribe to predicates arbitrary properties of objects in the universe of discourse then, as we have seen, the designata of predicates will not themselves generally be members of that universe. Hence, at least some entities to which predicates can make reference are such that no existential statements of either form can be true of them. In this sense it may be said that the realists in question ascribe to predicate entities

of a kind which do not, in general, "really exist."

Various positions of realism and nominalism have been distinguished just with respect to the character of designation relations which may be employed in defining truth for sentences of subject-predicate form. The consideration of quantificational contexts and of the values of variables in such contexts is extraneous to that distinction. For this reason, the position of realism at issue here, unlike the doctrine of Platonism characterized by Quine,¹³ can be identified without inspecting the existential assertions expressed in the object-language of a theory, and without settling the difficult question regarding the ontological commitment of a theory. We hope to have shown how some positions of realism and nominalism can plausibly be treated, not so much as doctrines about what there is, but rather as proposals about the manner in which various theories of reference might be constructed.

The University of Rochester

Received April 11, 1968

¹² Jaakko Hintikka, "Existential Presuppositions and Existential Commitments," *The Journal of Philosophy*, vol. 56 (1959), pp. 125-137.

¹³ W. V. Quine, *From A Logical Point of View*, *op. cit.*, pp. 127-128.

VIII. ON COMPLETENESS IN THE LOGIC OF QUESTIONS

DAVID HARRAH

I. INTRODUCTION

IN this paper we begin by considering the question: When is a logic of questions complete? Next we note the diagonal arguments of Cantor, which can be used to show that many logics of questions are incomplete. Then we consider the question, What is a question? Finally we outline a generalization on the idea of *question*, and correspondingly a wider framework for studying completeness.¹

Different theories of questions formalize the key notions in different ways. Accordingly, if we want to survey a wide range of theories, we should not try to formalize the key notions too closely or too early, lest we thereby exclude from consideration theories whose formalization differs from ours. At the start, at least, it is best to be content with the following simple and rough characterization of the question situation:

For a given question *q*, there are in addition to *q* the following items: a language *L*, zero or more presuppositions of *q*, zero or more answers to *q* (some of these being direct answers, some being indirect, some complete, some partial), and—if the foregoing entities are nonlinguistic—also the interrogative which expresses *q*, some sentences of *L* which express the presuppositions of *q*, and some sentences of *L* which express the answers to *q*. Exactly what the users of *L* want—i.e., what they want questions to be, what sort of questions they want, what they want to allow as answers, and what sort of asking-and-answering system they want—will vary from one situation to another. The questions and answers appropriate in library-use situations, e.g., differ from those appropriate in

courts of law, science, and psychotherapy. How questions and question systems can differ, and correspondingly how broad our notions of *question*, *system*, and the like, must be, will become evident below.

II. KEY PROPERTIES

Without considering particular areas of application, and the peculiarities of different logics which result from the peculiarities of different areas, we may consider certain properties which all logics of questions might have, or which all question logicians might be interested in. There does not yet exist a consensus on what properties a logic must have if it is to be a good logic of questions; to see this, one need only compare the different conditions of adequacy prescribed by different workers in this area.² It will be useful, however, to note here five kinds of property. These properties are important because of their interconnections; in particular, it is important that certain completeness properties are obtainable only if certain other properties are sacrificed. At this point we can give only a rough characterization of these properties; a more detailed and precise characterization can be given later, in terms of the formalization of the particular language involved.

(1) *Articulation Properties*. These are compounding and embedding properties. Suppose we have a logic in which the result of writing a declarative, followed by a sign for the material conditional, followed by an interrogative, is well-formed, perhaps to provide for asking a conditional question. In this case we could say that interrogatives articulate with declaratives. If the syntax is such that an interrogative cannot be

¹ Sections I–III of this paper incorporate material presented in talks given at Claremont in December 1966, at the University of Connecticut in February 1967, and in the Mahlon Powell Lecture Series at Indiana University in March 1967. I am grateful to all three audiences, and to Professor Nuel D. Belnap, Jr., for valuable comments and suggestions. Sections IV–VI extend and develop my “Erotetic Logistics,” forthcoming in Karel Lambert, (ed.) *The Logical Way of Doing Things: Philosophical Essays in Honor of Henry S. Leonard*. Research for this project was supported by the University of California Intramural Research Grant No. 7062.

² Compare, e.g., D. Harrah, *Communication: A Logical Model* (Cambridge, Massachusetts, MIT Press, 1963), N. D. Belnap, Jr., *An Analysis of Questions: Preliminary Report* (Santa Monica, Systems Development Corp., 1963), and Lennart Åqvist, *A New Approach to the Logical Theory of Interrogatives, Part I, Analysis* (Uppsala, Almqvist & Wiksell, 1965).

negated, e.g., then we could say that this system has only a partial articulation. Most of the logics of questions which have thus far been developed have no more than a partial articulation, but it seems to be an interesting problem to determine, for each logic of declaratives, whether a logic of questions can be added such that complete articulation is meaningful and possible.³

(2) *Consistency Properties.* If a logic of questions is built onto a logic of statements, the consistency properties of the latter determine consistency properties of the former. For example, if the replies to q are statements, or if the presuppositions of q are statements, then q can be said to have consistency properties as a function of the consistency properties of those statements. In some developments a question system has consistency properties which are peculiar to question logic. E.g., suppose that a logic allows biconditional questions. The questioner is guilty of some kind of inconsistency if he says both "If and only if F is the case, is it the case that G ?" (G here being a statement) and "If and only if F is not the case, is it the case that G ?"

(3) *Effectiveness Properties.* Let us identify the effective with the recursive and the constructive with the recursively enumerable. Suppose the questioner Q wants his questions to be mechanically recognizable as being questions, so that the respondent R can recognize mechanically when a question is being put to him and when a question is not; in this case Q will want *question* to be effective. Suppose on the other hand that what Q needs is not mechanical recognizability but long-run "positive-instance" reliability, where R is able to generate all questions and thus eventually recognize as questions all the things which happen to be questions; in this case Q 's needs will be satisfied by allowing *question* to be noneffective but requiring it to be constructive.

(4) *Completeness Properties.* The completeness properties which are relevant to the logic of questions include not only the completeness properties relevant to statement logics but also properties peculiar to question logic. To theorize about information-exchange situations, Q and R need the notion of *possible information exchange*, and want an explication of the notion of *all possible information*

exchanges. Note that, because Q has only a limited number of interests or only a limited ability to process information, what is at issue here is not completeness in the sense of being able to request all possible information, but completeness in the sense of being able to request any limited package of information. At a more abstract level, what Q and R want is an explication of *all possible answer requests.* To explicate such notions they form concepts of various completeness properties. These concepts can be defined in terms of such features of L as the number of sorts of things which can be asked about, the number of expressions of L available for use in asking questions (relative to the number of possible questions), the number of expressions available for use in answering questions, and the like.

Besides information exchange, there are other motivations which lead to completeness concepts. Suppose we develop a theory of questions which takes account of human limitations in various ways, in particular by distinguishing between a question's being *good* in some syntactical or semantic sense and its being *askable*, or between *having a true direct answer* and *being truly answerable*. It might be that q is expressible only by an interrogative which is too long to be communicated, or q might be communicable but have no true direct answer short enough to be communicated, or q might be truly answerable by one person but not by another, or q might call for answers which the questioner cannot accept (e.g., "What is a true sentence which I shall never believe?"), etc. Assuming that we could make precise the notions of *askable*, *answerable*, and the like, it would make sense to investigate whether a given logic of questions is complete with respect to askability, answerability, and the like; and in this area we should expect the incompleteness results to outnumber the completeness results.⁴

(5) *Coherence Properties.* Consider Frege's theory of meaning. For him, roughly speaking, the meaning of a compound is the corresponding compound of the meanings of the component parts. This suggests that there is an isomorphism between the syntax of an expression and the meaning of the expression, such that the users of the language can use syntax as a guide to meaning. Frege's theory was developed

³ H. Reichenbach, in his *Elements of Symbolic Logic* (New York, Macmillan, 1947), sect. 57, and R. B. Angell, in his *Reasoning and Logic* (New York, Appleton-Century-Crofts, 1964), sect. 6-2.1, state that questions cannot be negated. Åqvist, *op. cit.*, does not permit quantification into interrogatives.

⁴ I believe the first attempt to specify completeness concepts for a logic of questions is that of Belnap, *op. cit.* Belnap considers several important concepts, points out several completeness and incompleteness properties of his system, and (p. 136) cites N. Rescher's example of a question which is logically but not humanly answerable. The completeness of a calculus of interrogative operators, specified in terms of the set of all truths of the calculus, is noted in Tadeusz Kubiński, "O pojęciu pytania i rachunku operatorów pytajnych," *Ruch Filozoficzny*, Rok 1966, Tom XXV, p. 73.

for the logic of statements. Do we want something analogous, with principles of coherence based on isomorphism, for the logic of questions?

Suppose that a logic allows compound questions, such as conjunctions or disjunctions of questions. Then one coherence principle is that the presuppositions of a compound question are compounded in the appropriate way from the presuppositions of the component questions. Similarly, direct answers of a compound are appropriately compounded from direct answers of the components. Perhaps we should also secure coherence relative to equivalence relations, so that, e.g., questions which are equivalent in some other sense are also equivalent in regard to their presuppositions or in regard to their direct answers.⁵

III. INCOMPLETENESS

We must now take note of an argument which shows that many approaches to the logic of questions, including the classic and standard approaches, are in a certain fundamental sense incomplete. Let us assume that the language L has a finite alphabet, and that expressions of L are finite strings of letters of the alphabet, so that the expressions of L can be alphabetically ordered. Let us suppose that the set of questions of L is recursively enumerable. This would be the case if, e.g., each question is expressed by an interrogative of L , the interrogatives are specified via metalinguistic schemas, the set of these schemas is recursively enumerable, and, for each schema, the instances (i.e., interrogatives) subsumed by it are recursively enumerable. Suppose next that each question either has denumerably many direct answers or can be assigned denumerably many direct answers in a harmless way (e.g., by adding instances of some truth-functional contradiction). Suppose further that, given a question q , the direct answers to q are recursively enumerable; that is, they can be generated from q . Finally, for simplicity, we may suppose that the direct answers are sentences, or are expressed by sentences, and that *sentence of L* is recursive. Assuming all this, we can use Cantor's diagonal method to construct a set of sentences which is not the set of direct answers to any question in the initial enumeration of questions. In fact, we can construct indefinitely many such new sets, and we can make sure that they are in a

certain sense interesting. One way of performing the diagonal construction is the following:

We first choose some positive integer j . Then for the alphabetically first member of the new set we choose the alphabetically j -th sentence after the first direct answer of the first question, and for the alphabetically $(i+1)$ th member of the new set we choose the alphabetically j -th sentence after the first direct answer d of the $(i+1)$ th question such that d is alphabetically later than the i -th member of the new set. Clearly, every question in the initial enumeration has some direct answer which is not included in the new set. Since j can be any integer from 1 onward, we can by this procedure construct indefinitely many new sets. Now we could also specify in advance some recursive property P —for example, the property of not being a tautology or the negation of a tautology—and specify that the construction is to move at least j sentences at each jump and is to keep moving until it finds a new sentence which has the property P . It is in this sense that our construction can produce new sets which are interesting.

One last step is needed before we can interpret these Cantorian results as showing that some question systems are incomplete. We must assume not only that each question has a set of direct answers, but also that to each set of sentences there corresponds a question for which the set is the set of direct answers. One might declare that his initial enumeration exhausted all questions, that Cantor's procedure does reveal a set of sentences which is not the set of direct answers for any question, and that we should take pity on such sets but that we need not scorn the initial enumeration for being incomplete. If on the other hand we believe that all sets of sentences are created equal, and that each one has the inalienable right to be represented by a question, then we must regard any constructive census of questions as incomplete. To honor the unlisted questions we should perhaps follow the example of number theory and say that the listed questions are "rational" but the unlisted ones are the "real" questions.

Notice the range of approaches affected by this argument. The argument is neutral regarding the syntax and semantics of interrogatives. The argument applies to some nonstandard approaches, including some which use languages which are infinitary in various respects. Notice also that

⁵ For a discussion of a different but related concept of *coherence*, see N. D. Belnap, Jr., "Questions, Answers, and Presuppositions," *The Journal of Philosophy*, vol. 63 (1966), pp. 609-611.

similar methods can be used to show that many logics of questions are incomplete at other points or in other respects.⁶

At this stage the conclusion emerges that, roughly speaking, if one uses a standard language and wants every set of sentences to be represented by a question, then: (1) If *question* and *direct answer* are constructive, the logic is incomplete. (2) If *direct answer* is constructive and one wants completeness, then *question* must be nonconstructive; or, if *direct answer* is constructive, then one must choose between constructivity and completeness. Some logicians have already in effect opted for completeness, while others have insisted on constructivity.⁷ As yet, no one seems to have developed a system in which *question* is constructive and *direct answer* is not; perhaps this shows that the most important feature of a question system is neither completeness nor constructivity of *question*, but rather the constructivity of *direct answer*; perhaps it shows that, although all men by nature desire to know, they tend to sympathize with the respondent rather than with the questioner.

One worthwhile enterprise would be to explore the consequences of insisting on constructivity, and to follow the program of Belnap and Åqvist: for each language L , to determine what types of question can be defined for L . Another worthwhile enterprise would be to follow a program which seems to be a natural development on the approach of Stahl: for each (sufficiently rich) language L , to develop means within L of describing arbitrary sets of expressions, so that the user of L can ask not only his initially provided rational questions but also his newly discovered real questions.

Meanwhile, another worthwhile enterprise is to re-examine the concept of *question*. The Cantorian argument forces us to do this in any case, because the Cantorian results are interpretable as incompleteness results only if we adopt a certain concept of *question*. We might look for an argument which establishes one of the already existing concepts as the definitive concept. Alternatively we might try to establish *question* as a "family resemblance" concept. A third possibility is to look for a notion which is deeper than the notion of *question*, which will yield as special cases all the special notions of *question* which have hitherto been of importance, including both the notions required for the constructive approaches and those required for the

nonconstructive. The remainder of this paper is devoted to this third possibility.

IV. A CONCEPT OF INTEREST

The basic parameter in our analysis is the language L , and for the analysis we must make certain assumptions about L . The definitions presented below presuppose very few assumptions about L (and thus our analysis applies to a very wide class of languages), but it will simplify our exposition if we assume now that L is a formal language whose logical apparatus includes the first-order predicate calculus with identity. We want to allow L to contain much more than this, of course, and some of the examples presented below will indicate various kinds of additional content that L might have. In particular we allow L to have complex terms such as definite descriptions and indefinite descriptions (the latter corresponding to phrases like "a number between 4 and 9"), so that variables can occur either free or bound inside a term. Let us assume that *expression*, *term*, *wff*, *sentence*, *derivation*, and *implies* have been defined for L , and that the semantics of L has been developed to the point where *interpretation* of L (or, *model* of L , in the usual model-theoretic sense) has been defined. For convenience we allow the empty set to count as a sequence of expressions. Then, with *expression*, *sentence*, *interpretation*, and the *ke*, relativized to L , we define:

D1: A *semantic pair* is an ordered pair $\langle Y, Z \rangle$ such that Y is a set of interpretations and Z is a sequence of expressions.

D2: A set S *logically contains* an expression X , *relative to* a set S' of wffs, if and only if either:

- (1) X is a wff and there is a subset W of S such that every member of W is a wff and X is derivable from the union of W and S' ; or
- (2) there are terms t and t' and an individual variable x (not free in t or t') such that t is in S and either
 - (a) X is a wff and X is derivable from S' together with $(x=t)$, or
 - (b) X is t' , and $(x=t')$ is derivable from S' together with $(x=t)$.

D3: An *interest*, I , is an ordered quintuple $\langle Intp, Bk, Tp, Sat, Prt \rangle$

such that:

- (1) *Intp* (which we call the "interpretation set" of I) is a non-empty set of interpretations.

⁶ See, e.g., Belnap, *An Analysis of Questions*, op. cit., p. 74, on category predicates.

⁷ Stahl and Belnap, respectively. See Belnap's *Analysis*, op. cit., and my review of Stahl in *The Journal of Symbolic Logic*, vol. 28 (1963), p. 259.

- (2) *Bk* (the "background" of *I*) is a set of sentences.
- (3) *Tp* (the "topic" or "area" of *I*) is a non-empty set of expressions.
- (4) *Sat* (the "satisfaction set" of *I*) is a set of semantic pairs $\langle Y, Z \rangle$ such that $Y = \text{Intp}$ and each member of Z is a member of *Tp*.
- (5) *Prt* (the "partial satisfaction set" of *I*) is a set of semantic pairs $\langle Y, Z \rangle$ such that Y is a subset of *Intp*.
- (6) If $\langle Y, Z \rangle$ is in *Prt* and Z is non-empty, then there is a $\langle Y', Z' \rangle$ in *Sat* such that Z' is non-empty and each member of Z is logically contained in the union of *Bk* and the set of members of Z' .

D4: An *interest contract*, *IC*, is an ordered triple $\langle I, K, Pf \rangle$

such that:

- (1) *I* is an interest.
- (2) *K* is a set of pairs $\langle Y, F \rangle$ such that Y is a non-empty set of interpretations and F is a sentence.
- (3) *Pf* is a function whose domain is a set of semantic pairs.

D5: An *interest* (or, *interest contract*) *expression function* is a function whose domain is a set of interests (or, interest contracts) and whose range is a set of semantic pairs. (If the function takes *I* as argument and assigns $\langle Y, Z \rangle$ to it, we say that $\langle Y, Z \rangle$ expresses *I* with respect to the function.)

D6: An interest (or, interest contract) *expression function* is *finitary* if and only if, for every $\langle Y, Z \rangle$ in the range of the function, Z consists of just one member; it is *quasi-finitary* if and only if, for every $\langle Y, Z \rangle$ in its range, Z is finite; and it is *non-finitary* if and only if, for some $\langle Y, Z \rangle$ in its range, Z is infinite.

What we are doing here is generalizing from the standard conception of the *question* situation to a conception of the *interest* situation. In the standard question situation the language *L* is given and one interpretation of *L* is adopted. The notion of *question* is defined, or various types of question are defined; and for each type of question *direct answer*, *partial answer*, and *presupposition* are defined. The questioner has a body of background knowledge; he agrees that, when he asks a question, he is thereby committed to the presuppositions of the question; and he promises that he will pay for answers to his questions at a certain rate specified in advance.⁸

In the interest situation, now, the language *L* is fixed and one interpretation is specified as the

intended or normal interpretation, but the unit of communication is not an expression of *L* by itself (to be interpreted in the normal way) but a semantic pair. The message unit is a sequence of expressions, together with an "interpretation indicator," which is a linguistic or nonlinguistic device for indicating which interpretations of *L* are to be used in interpreting the expressions. The interpretation indicators, of course, must themselves be interpreted by a code fixed in advance. (Note that natural language systems include devices which serve as interpretation indicators; typical devices of this sort are intonation, facial expression, the "Once upon a time" phrase, and "scare" quotes.)

The communicator has a general area or topic of interest, and a background perspective on it (cf. "I have a geologist's interest in Lake Henshaw"). By means of some expressions of *L* he describes or indicates what sort of thing will count as directly satisfying his interest, and what sort of thing will count as partially satisfying it. He also expresses what he commits himself to in the way of presuppositions (namely, the entities in *K*), and what payments he will make (via the function *Pf*) to the respondent for various replies the respondent might give him.

V. APPLICATIONS

The interest situation includes the normal interrogative situation as a special case, as follows: Let the interest *I* be $\langle \text{Intp}, Bk, Tp, Sat, Prt \rangle$. Let the interest contract *IC* be $\langle I, K, Pf \rangle$, where *K* consists of just a single pair $\langle Y, F \rangle$. Let the Y here and in the $\langle Y, Z \rangle$'s by which *I* and *IC* are expressed be the same as *Intp*; let *Intp* consist of just the intended interpretation of *L*; also, let these Z 's be finite. Suppose that *Bk* is the communicator's knowledge, that *Tp* consists of sentences, and that both *Sat* and *Prt* consist of pairs $\langle Y, Z \rangle$ such that Z consists of just one sentence. Suppose that the *F* in *K* is a sentence which is true in the intended interpretation just in case there exists a Z such that $\langle Y, Z \rangle$ is in *Sat* and all the members of Z are true in the intended interpretation. Finally, suppose that *Pf* is some semantic information function, possibly vacuous. Then, in effect, *IC* is a question, *Sat* is the set of its direct answers, *Prt* is the set of its partial answers, and *K* contains its principal erotetic presupposition.⁹

For questions, Hamblin has postulated that

⁸ See my *Communication*, *op. cit.*, sects. 6-10.

⁹ Concerning this sort of presupposition, see Belnap, "Questions, Answers, and Presuppositions," *op. cit.* Concerning *Pf* functions, see my *Communication*, *op. cit.*, sect. 11.

knowing what counts as an answer is equivalent to knowing the question.¹⁰ A generalized version of this principle holds for interests, in the following sense: The *Bk* set functions only to determine *Prt*, and the *Tp* set functions only to determine *Sat*. Insofar as *Sat* is "equivalent" to *Tp* (e.g., in the sense that, for every member *F* of *Tp*, there is a $\langle Y, Z \rangle$ in *Sat* such that *F* belongs to *Z*), then knowing the interest is equivalent to knowing what satisfies the interest and what partially satisfies the interest. On the other hand, of course, for an interest where *Sat* does not exhaust *Tp*, or is not equivalent to it in some other sense, knowing what satisfies the interest is not equivalent to knowing the interest.

Notice that we have made only minimal specifications concerning the *Z*'s by which *I* and *IC* are expressed. Our definitions allow *I* and *IC* to be expressed simultaneously by the same $\langle Y, Z \rangle$, and allow *Z* here to consist of a single expression, which may be an interrogative, an imperative, a declarative, or a sentence of some other kind. The expressions in *Z* may have any kind of semantics; they may be interpreted as making assertions either directly (as in "I want to know the name of that substance") or indirectly (as in "Why does that substance contract when heated?"). Putting it more generally, our definitions place no constraints on the coherence properties of the *I* and *IC* expression functions.

Notice also our generality in other respects. Our definitions make no specifications concerning effectiveness. Given *Z*, and given that $\langle Y, Z \rangle$ expresses *I* or *IC*, it might or might not be the case that *Sat* is recursive, or recursively enumerable. Thus in particular we allow *direct answer* to be non-effective or non-constructive. We allow *Sat* or *Prt* to be empty; we allow *Sat* and *Prt* to include a $\langle Y, Z \rangle$ such that *Z* is empty (constituting a null reply) or such that *Z* contains all the members of *Tp*. We allow *Tp* to contain terms as well as formulas, and *Sat* to contain $\langle Y, Z \rangle$ such that *Z* is a sequence of terms, so that the user of the system can say "Give me a list of names such that . . ."

We allow *Y* to contain an interpretation of *L* which is not the intended interpretation of *L*, so that the communicator can be "interested in" things in the universe of *L* which are not named under the intended interpretation but are named under the alternative interpretation in *Y*. Now if *L* included set theory and a descriptive operator, some things

not named under the intended interpretation might still be adequately describable, so that the *Y* parameter would not be needed. The same point holds for the case where the communicator is interested in certain properties which are not expressed by predicates of *L* under the intended interpretation. It seems likely that the *Y* parameter will show its usefulness when we come to situations where the communicator wants to formulate his interests in terms of infinite sets of interpretations. (Perhaps this is the case with "I am interested in all sets whose defining characteristic has the form . . .") To establish the need for *Y* and exploit the power of *Y* are matters for further study. Naturally we should first specify *L* and then develop the theory of interpretations of *L*. After that we should see what sort of system can be developed whereby particular interpretations and classes of interpretations can be described (or named, or somehow indicated) by expressions of *L*. We could then let some of the pairs $\langle Y, Z \rangle$ above, where *Y* is a set of interpretations, be associated with pairs $\langle Y^*, Z \rangle$ such that *Y*^{*} is an expression, or sequence of expressions, which describes (or names, or indicates) *Y*; the pairs $\langle Y^*, Z \rangle$ could then be taken as expressing the pairs $\langle Y, Z \rangle$.

Finally, note the power that results from allowing *Z*, in the various pairs $\langle Y, Z \rangle$ above, to be infinite. Suppose, e.g., the communicator wants the respondent to choose a true sentence from an infinite set *S* of sentences, but there is no (finite) expression *F* agreed on by communicator and respondent as expressing *S* in the sense that the members of *S* can be generated from *F*; suppose, in other words, that the only effective way to present *S* is to present its members one after another. On the other hand, even where there is a finite expression which can convey the question, the direct answers might themselves be infinite, as in "What are all the prime numbers?" or "What are all the truths of number theory?" It is to make possible the presentation of infinite questions, or infinite answers, or infinite lists of any kind, that we allow *Z* to be infinite. It might seem that this system favors questioner *Q* over respondent *R*, in that in many cases *Q* need not complete his infinite list (for after a finite time he will have presented an expression which *R* immediately recognizes as "solving the problem"), while *R* does have to complete his infinite answer if the answer is indeed infinite. On second thought, however, we see that it is *Q* who

¹⁰ C. L. Hamblin, "Questions," *Australasian Journal of Philosophy*, vol. 36 (1958), p. 162.

suffers, for he has committed himself to desiring the impossible, and *R* can be paid for at least a partial answer.

VI. COMPLETENESS?

For the interest situation, what sorts of completeness are attainable? It is, of course, impossible to map all the interests *I* or interest contracts *IC* into the expressions of *L* in such a way that each expression of *L* expresses at most one *I* or *IC*; so the classical (finitary) kind of expressive completeness is unattainable. But in the interest system as determined by the definitions above, expression is by means of semantic pairs $\langle Y, Z \rangle$, where *Z* is a sequence of expressions, so our problem in effect is to map *I*'s and *IC*'s into the set of sequences of expressions of *L*.

Now obviously we can obtain completeness with

respect to the *Bk* and *Tp* sets; moreover, because we can express them directly via sequences, we can preserve consistency and coherence. The question of whether we can similarly secure completeness with respect to the *Intp*, *Sat*, and *Prt* sets hinges, on the one hand, on what content we give *L* and, on the other, on what content we give ourselves in the meta-level set theory by which we determine the variety of *Intp*, *Sat*, and *Prt* sets. We can simplify things by allowing *L* to include some set theory and syntax, so that expressions of *L* can describe sets of expressions of *L*. The major questions then would be: For any given expression system, how interesting are the *I*'s and *IC*'s which are left unexpressed in that system? Or, what restrictions on the nature of *Intp*, *Y*, *Sat*, and *Prt*—sufficient to insure completeness of expression—might reasonably be imposed? To explore these matters in detail is a challenge to further inquiry.

University of California, Riverside

Received March 18, 1968

IX. ART, ARTIFACTS, AND INTENTIONS

MARCIA M. EATON

EVERY reader of literary criticism knows that there is a great deal of talk about the role which the intentions of authors of certain works play in literary explanation. Critics often say that an author meant such and such by some word or phrase, that he wanted to do something or other in one of his works. However, in recent years, particularly since W. K. Wimsatt and Monroe Beardsley's discussion of the matter,¹ the legitimacy of reference to an author's intentions has been questioned. Many critics now take the view that talk about intentions is out of place in the discussion of literature. Several other critics, however, have remained unconvinced and for various reasons continue to consider reference to intentions a valid procedure.

The main reason that the problem of intention is still a problem is due, I believe, to a failure to understand the nature of a literary work. Once we understand something about that nature, namely that a work of literature is a linguistic object, the role of intentions becomes more obvious. I wish first to show how so viewing a literary work solves one of the more specific problems nested within the intentional controversy and then sketch a way in which it can shed light on more general problems.

For some time aestheticians have been bothered with the question of whether or not a thing must be an artifact for it to count as a work of art. Can paint drippings of which the artist is unaware properly be called a painting? Can a monkey's random typewriter pickings, if they result in a syntactically and semantically meaningful arrangement of words, be correctly termed a novel? Is a piece of driftwood untouched by human hands ever a piece of sculpture?

George Dickie maintains that one necessary condition of a thing's being a work of art is artifactuality.² I believe that he is correct. Perhaps it is not enough that a thing be an artifact in order for it to be called a work of art. Nonetheless, it is not possible that a thing count as a work of art unless it is also an artifact.

Other philosophers feel that this is not right. I use

"feel" here quite intentionally—for I believe that it is little more than a feeling. These philosophers argue that the fact that someone might point to a piece of driftwood and say, "There is a work of art," and further say, "That is not an artifact," proves that artifactuality is not a necessary condition of a thing's being a work of art. But I think that there are good and decisive reasons for not agreeing with this view.

It is certainly true that philosophers who engage in linguistic analysis come down hard on what people say. They do so for philosophic reasons. It is a natural result of the belief that the way people talk about the world is more illuminating than the world itself. Finding one person who says that a thing is not such and such can be important. It may even be necessary that it be brought under consideration by a linguistic analyst who says that a thing *is* such and such. But even a linguistic analyst is willing to admit that some people say incorrect things. That is, merely being able to find a person who talks differently does not constitute proof that a suggested analysis is wrong.

Suppose I am walking on the beach with a friend and we come upon a dead whale that has washed onto the shore. My friend may say, "There is a dead sea gull." My friend may need glasses. He may not speak English. He may not be talking about the object at which I am looking. All of these are possible explanations of his making the statement, "There is a dead sea gull." However, we would not say that my friend had proved that I do not understand whalehood. When my friend points to the dead whale and says, "There is a dead sea gull," he is simply saying something incorrect. He is just wrong. He has certainly not proved that what I heretofore believed to be necessary and sufficient conditions of being a dead whale are not such conditions.

Now suppose we walk a little further down the beach and come across a piece of driftwood and my friend says, "There is an artifact." I would say exactly the things I said about his saying, "There is a dead sea gull." He is wrong. I may question him and

¹ W. K. Wimsatt, Jr., *The Verbal Icon* (Lexington, Ky., 1954) and Monroe Beardsley, *Aesthetics* (New York, 1958).

² This view of Dickie's will be elaborated in a forthcoming paper of his.

say, "Do you think that it has been made by a human or that it has been modified in some way by a human?" He may answer, "No, but I still say that it is an artifact." Then I am sure that he has just said something that is false. I do not believe that he has shown me that human workmanship or modification are not necessary and sufficient conditions of artifactuality. (This "analysis" of "artifact" is due not to me but to Webster.)

But suppose that my friend points to the driftwood and says, "There is a work of art." He agrees that it is not an artifact but is a natural object, but still insists that it ought to be called a work of art (still in the descriptive, not evaluative sense). Dissension with regard to whales and artifacts is possible but not disturbing. However, there has been a significant contingent of philosophers of art which maintains that calling the driftwood a work of art descriptively, is not as strange, not as obviously an error as calling the whale a sea gull or the wood an artifact is an error. In fact, there have been those who have said that it is not an error at all, that it is in no way strange to call a piece of driftwood a work of art. What I hope to show is that it is an error, and that, even if this does not sound strange, it nonetheless should.

Suppose my friend precedes me a few miles down the beach. He hears on his portable radio that a tidal wave will hit in one hour. He wants to warn me of this fact, but he also feels obligated to warn other people on down the beach. He does not have time to do both. So he hastily gathers up several pieces of driftwood and spells out "TIDAL WAVE IN ONE HOUR," and hurries on down the beach to publicize the impending danger. One correct description of what my friend has done is that he has used pieces of driftwood to warn me of the approaching tidal wave.

Now let us alter the situation somewhat. Again suppose my friend is preceding me on the beach and hears over his portable radio that a tidal wave will strike in one hour. Suppose further (this may require quite a stretching of imaginations, but is still possible) that he looks ahead of himself and sees that wind and waves have arranged several pieces of driftwood in the configuration TIDAL WAVE IN ONE HOUR. My friend leaves it untouched and hurries on down the beach to warn the other people. Again, one correct description of what my friend has done is that he has used pieces of driftwood to warn me of the impending tidal wave.

Prior to my friend's use of them as such, the pieces of driftwood arranged by wind and wave in

the configuration TIDAL WAVE IN ONE HOUR do not constitute a warning. This is just because a necessary condition of an arrangement of words being a warning is that it be the result of a human intention. A spoken or written string of sounds or shapes does not count as a warning unless it is intended that it so count. My friend intends that I take the arrangement of driftwood as a warning. Unless he (or someone else) does so intend, it cannot be a warning—it is nothing more than a strange coincidence. Likewise, before my friend uses it as a warning, the arrangement of driftwood is not an artifact.

Suppose my pet monkey sits at my typewriter and picks at certain keys. When he is through picking, I discover that he has typed out a group of words arranged grammatically and meaningfully in a regular metrical and rhyme scheme. I quickly take the paper out of the machine, put it in a stamped envelope, and mail it off to the *Atlantic Monthly*, with an accompanying note which says, "Here is a poem written by my monkey. Please publish it." The editor who receives that letter would probably throw it and the poem into the wastebasket, dismissing it as a hoax. Even if he did not so dismiss it, even if he believed that what he held in his hand had many of the good-making characteristics of poetry, he would still, undoubtedly, be reluctant to publish it. His hesitation is appropriate. He believes that monkeys are not poets. He would feel the same reluctance if I had instead told him that a gust of wind wrote the poem by arranging some words and punctuation marks I had on my desk in a poem-like way.

The hesitation to accept the monkey-poem and wind-poem as full-fledged examples of poetry is due not just to the improbability of such happenings. It is due to a logical point. Just as there can be no warnings independent of intention, so there can be no poems independent of intention.

Ordinarily the act of warning takes the form of uttering or writing certain words. However, in the example of the driftwood already arranged on the beach, warning takes the form of using already existing shapes. Ordinarily writing poetry takes the form of writing words "from scratch" (excuse the pun). But in the case of the monkey and the wind, writing poetry takes the form of using already existing shapes. Merely arranging shapes or words in a poem-like form does not make the monkey or the wind a poet. Nor are those shapes artifacts. It is my using them in such and such a way, e.g., mailing them to the *Atlantic Monthly* which makes me a poet,

and a maker of artifacts. Putting things into stamped envelopes and mailing them is admittedly an extraordinary way of becoming a poet. However, it is still a possible way of so becoming.

Now let us consider how the artifactuality issue is connected with the problem of intention in literary explanation. An arrangement of words can be neither a warning nor a poem unless someone intends that it be such. This is true of all objects which are the result or product of linguistic actions.

There are several different types of linguistic actions which an author may perform. As a basis I use Austin's classification. An author performs a locutionary act when he writes down words, sentences, lines of poetry, etc. When a writer makes assertions, asks questions, issues commands, etc., he is performing illocutionary acts. When a writer produces certain effects on the part of his readers such as informing, persuading, deceiving, etc., he is performing perlocutionary acts. All of these types of linguistic actions are continually being performed by ordinary (nonliterary) users of language. Writers of literature commonly (though not always) perform other types of linguistic actions—actions which may be performed by ordinary language-users, but most often are not. First, an author frequently performs what I call translocutionary acts, whereby he attributes or transfers illocutions to dramatic speakers or protagonists. (When Hamlet says, "Get thee to a nunnery," a command is being made, not by Shakespeare, but by a dramatic speaker, viz., Hamlet.) Secondly, an author may use language figuratively, an action different from any of the above types.

In addition to these linguistic actions an author may also perform nonlinguistic actions—actions which can be carried out without the explicit use of language, e.g., making money, praising God, expressing melancholy.

Since a work of literature is a linguistic object, it is to be explained in very much the same manner that we explain other kinds of linguistic objects. Unfortunately the term "explanation" when employed in connection with linguistic objects, particularly literary linguistic objects, is extremely vague. It is for this reason that I wish to divide the activity of explaining into two different sub-activities: explication and interpretation. I am not offering an analysis of either of these terms. Rather I wish to suggest a way of using them in discussions of literature which is stricter than their use heretofore.

The difference between the explication of a

literary work and the interpretation of a literary work is best understood in terms of the distinction between a linguistic object and a linguistic action. Quite simply, a linguistic object is explicated; a linguistic action is interpreted. Again quite simply, a linguistic object is explicated by giving the meanings of the words of which it consists. A linguistic action is interpreted by describing what is being done, primarily in terms of the way or ways in which certain linguistic objects are consciously being used.

I may understand the sentence a person writes or utters without understanding the linguistic action which he is thereby performing. This is because understanding the latter requires an interpretation as well as an explication of the words which are spoken or written. Suppose Jones says, "This suitcase is heavy." I may give you a complete explication of that sentence by giving you the meanings of all the words employed. Here by "meaning of words" I refer to the kind of information one would get from a dictionary. Thus the explication may very well include statements concerning the *typical* use of the sentence "This suitcase is heavy." However, it is quite possible that given such an explication you are still unable to interpret Jones's action, for it is a fact about language that sentences can be used in non-typical ways. Words having the same *meaning* can be *used* in different ways. I give you an interpretation of Jones's action when I further tell you that Jones was making an assertion, or requesting assistance in carrying the suitcase, or telling a lie, or voicing a complaint, or joking, or praising my packing job, etc.

This difference between an explication and an interpretation carries over to literary works. As a linguistic object, a poem, for example, is explicated. As the record of the linguistic and translinguistic actions of a writer and as the linguistic actions of a dramatic speaker, a literary work is interpreted. Anyone who has ever read a "difficult poem" is painfully aware of the fact that it is often possible to explicate all of the words used in a poem without being able to give a satisfactory interpretation of it. Consider the opening line of "The Waste Land," "April is the cruellest month." The meanings of the words used here offer no difficulty. There is a clear sense in which everyone who understands English understands the line. But the interpretation of it as a line of poetry requires more than simply knowing the meanings of the words involved. The explication of a work of literature is rarely the subject of controversy; the interpretation often is.

As meanings of words exist necessarily independent of the particular intentions of language users, intentions play no role in explication. In ordinary discourse I cannot arbitrarily intend the words I utter to have meanings which they do not have. Writers of literature work within the framework of a language—an already existing language. The words they use have meanings, and writers are no more at liberty to intend the words they use to have meanings different from the ones which they do have than is the nonliterary writer or speaker. It is true that writers (e.g., James Joyce) often “play” on words in all sorts of ways. But even this play demands that those words have their core meaning. When I say “Mabel is a vixen,” I am not literally identifying her as a member of the class of vixens. Neither am I changing the meaning of the word “vixen.” Rather I am describing Mabel figuratively; and the description is effective precisely because the core meaning of “vixen” remains unchanged. The particular intentions which a person has before or during his use of a word cannot radically affect the meaning which is already associated with that word. It is for this reason that the explication of a linguistic object in general, and of a literary work in particular, does not involve a reference to the intentions of the speaker or writer.

Intentions (particular intentions of particular individuals) enter rather at the level of interpretation. It will be recalled that we interpret not linguistic objects per se, but the linguistic actions which make use of those linguistic objects. The interpretation of a linguistic action consists of a description of what a person is doing, i.e., what he is doing above and beyond the simply writing or uttering of words. When Jones said, “This suitcase is heavy,” I give you an interpretation of his action when I tell you, for example, that he is requesting my assistance. My ability to distinguish between Jones’s simply asserting that the suitcase is heavy and requesting my assistance, actions which can both be performed by uttering the words “That suitcase is heavy,” depends upon my knowing what Jones’s intentions in uttering those words are. The interpretation of this linguistic action performed by Jones consists largely of this kind of distinguishing, based on a knowledge of Jones’s intentions. It is not going at all far afield to say that in everyday conversation we are constantly making interpretations of this sort.

It is my contention that the interpretation of literary works qua linguistic actions (i.e., as a record of a writer’s linguistic actions or as uttered by a

dramatic speaker) is not essentially different from the interpretations which we make when we attempt to understand the speeches which we hear in everyday, nonliterary conversation. It is precisely for this reason that the intentions are, or can be, of extreme importance in the interpretation of literature. Consider this poem: “God is great, God is good, God is just like Robin Hood.” An explication of that poem consists in giving the meanings of the words in that poem. An interpretation consists in telling how the poet used those words, or what he intended to do when he wrote them, e.g., be sarcastic, write a funny poem, write a beautiful poem, praise God, praise Robin Hood, etc.

Intentions are related to language and the use of language in a very fundamental way. It is intentions which turn linguistic objects into linguistic actions. Thus, knowledge of intentions naturally aids us in interpreting linguistic actions. We may not care what an author’s intentions were. However, most people most of the time engage in the activity of reading—an activity which involves interpreting statements of a given person, viz., the author. We allow the interpretation we give to a work to differ from an author’s (where it is available) only with very careful consideration. Why this is so is related to the reasons people have for reading literature in the first place. There are, of course, all sorts of reasons for doing this, e.g., to learn about the human situation, to learn about something specific, e.g., the life of a doctor, the court of King Charles II, gamekeeping, etc., to increase knowledge in general, to have vicarious pleasures, to widen experience, to pass the time, and so on. At any rate, the enjoyment gleaned from reading works of literature per se, that is, as the work of a particular individual, demands a certain amount of truthfulness and faithfulness—a certain responsibility to that individual.

We could engage in an activity different from reading as such. We could stop talking about Milton’s poems and Shakespeare’s plays, and Hemingway’s novels, and discuss only individual reactions to works of literature as physical objects completely separated from their cause. Where the problem is crucial is in interpreting a work itself. We do speak of Milton’s poems. Insofar as Milton’s poems are his, they are so because of certain linguistic and nonlinguistic intentions which he had. We may not be able to spell out in any great detail what all or even most of those intentions were. But this is not to say that discovering or knowing what any given writer’s intentions were is not an important and legitimate part of interpretation.

We do, of course, have the problem that statements about the intentions of an author may not be available. In addition, there are problems even if those intentions are available and we want to use them in our interpretation. They may be irrelevant. The statements available may not constitute good evidence for knowing what an author's intentions

actually were. They may have been made, for example, by an incorrigible liar. But these are problems which are common to interpretation in general—not unique to literary interpretation. Evidence must always be examined carefully. However, good evidence should always be utilized.

University of Illinois at Chicago Circle

Received January 17, 1968

X. THE IRRELEVANCE TO RELIGION OF PHILOSOPHIC PROOFS FOR THE EXISTENCE OF GOD

STEVEN M. CAHN

PHILOSOPHIC proofs for the existence of God have a long and distinguished history. Almost every major Western philosopher has been seriously concerned with defending or refuting such proofs. Furthermore, many contemporary philosophers have exhibited keen interest in such proofs. A survey of the philosophical literature of the past decade reveals quite a concentration of work in this area.¹

One might expect that religious believers would be vitally interested in discussions of this subject. One might suppose that when a proof of God's existence is presented and eloquently defended, believers would be most enthusiastic, and that when a proof is attacked and persuasively refuted, believers would be seriously disappointed. But this is not at all the case. Religious believers seem remarkably uninterested in philosophic proofs for the existence of God. They seem to consider discussion of such proofs as a sort of intellectual game which has no relevance to religious belief or activity. And this view is shared by proponents of both supernaturalist and naturalist varieties of religion. For example, Søren Kierkegaard, a foremost proponent of supernaturalist religion, remarked: "Whoever therefore attempts to demonstrate the existence of God . . . [is] an excellent subject for a comedy of the higher lunacy!"² The same essential point is made in a somewhat less flamboyant manner by Mordecai M. Kaplan, a foremost proponent of naturalist religion, who remarks that the "immense amount of mental effort to prove the existence of God . . . was in vain, since unbelievers seldom become believers as a result of logical arguments."³

In what follows, I wish to explain just why religious believers have so little interest in philo-

sophic proofs for the existence of God. I wish to show that their lack of interest is entirely reasonable, and that whatever the philosophic relevance of such proofs, they have little or no relevance to religion.

The three classic proofs for the existence of God are the ontological, the cosmological, and the teleological. Each of these proofs is intended to prove something different. The ontological argument is intended to prove the existence (or necessary existence) of the most perfect conceivable Being. The cosmological argument is intended to prove the existence of a necessary Being who is the Prime Mover or First Cause of the universe. The teleological argument is intended to prove the existence of an all-good designer and creator of the universe.

Suppose we assume, contrary to what most philosophers, I among them, believe, that all of these proofs are valid. Let us grant the necessary existence (whatever that might mean) of the most perfect conceivable Being, a Being who is all-good and is the designer and creator of the universe. What implications can be drawn from this fact which would be of relevance to human life? In other words, what difference would it make in men's lives if God existed?⁴

Perhaps some men would feel more secure in the knowledge that the universe had been planned by an all-good Being. Others, perhaps, would feel insecure, realizing the extent to which their very existence depended upon the will of this Being. In any case, most men, either out of fear or respect, would wish to act in accordance with the moral code advocated by this Being.

Note, however, that the proofs for the existence of God provide us with no hint whatever as to which

¹ For a partial bibliography, see Robert C. Coburn's "Recent Work in Metaphysics," *American Philosophical Quarterly*, vol. 1 (1964), pp. 218-220. Two comprehensive treatments of the subject are Wallace I. Matson's *The Existence of God* (Ithaca, Cornell University Press, 1966) and Antony Flew's *God and Philosophy* (London, Hutchinson & Co., 1966).

² *Philosophical Fragments*, tr. by David F. Swenson (Princeton, Princeton University Press, 1936), ch. III, p. 34.

³ *The Future of the American Jew* (New York, The Macmillan Company, 1948), p. 171.

⁴ I am not concerned here with the implications of God's omniscience and omnipotence for man's free will. It is possible to interpret these divine attributes in such a way as not to entail the loss of man's free will, and for the purposes of this essay, I shall assume such an interpretation.

actions God wishes us to perform, or what we ought to do so as to please or obey Him. We may affirm that God is all-good and yet have no way of knowing what the highest moral standards are. All we may be sure of is that whatever these standards may be, God always acts in accordance with them. One might assume that God would have implanted the correct moral standards in men's minds, but this seems doubtful in view of the wide variance in men's moral standards. Which of these numerous standards, if any, is the correct one is not known, and no appeal to a proof for the existence of God will cast the least light upon the matter.

For example, assuming that it can be proven that God exists, is murder immoral? One might argue that since God created man, it is immoral to murder, since it is immoral to destroy what God in His infinite wisdom and goodness has created. This argument, however, fails on several grounds. First, if God created man, He also created germs, viruses, disease-carrying rats, and man-eating sharks. Does it follow from the fact that God created these things that they ought not to be eliminated? Secondly, if God arranged for men to live, He also arranged for men to die. Does it follow from this that by committing murder we are assisting the work of God? Thirdly, if God created man, He provided him with the mental and physical capacity to commit murder. Does it follow from this that God wishes men to commit murder? Clearly, the attempt to deduce moral precepts from the fact of God's existence is but another case of trying to do what Hume long ago pointed out to be logically impossible, viz., the deduction of normative judgments from factual premisses. No such deduction is valid, and, thus, any moral principle is consistent with the existence of God.

The fact that the proofs of God's existence afford no means of distinguishing good from evil has the consequence that no man can be sure of how to obey God and do what is best in His eyes. One may hope that his actions are in accord with God's standards, but no test is available to check on this. Some seemingly good men suffer great ills, and some seemingly evil men achieve great happiness. Perhaps in a future life these things are rectified, but we have no way of ascertaining which men are ultimately rewarded and which are ultimately punished.

One can imagine that if a group of men believed in God's existence, they would be most anxious to learn His will, and consequently, they would tend to rely upon those individuals who claimed to know

the will of God. Diviners, seers, and priests would be in a position of great influence. No doubt competition between them would be severe, for no man could be sure which of these oracles to believe. Assuming that God made no effort to reveal His will by granting one of these oracles truly superhuman powers (though, naturally, each oracle would claim that he possessed such powers), no man could distinguish the genuine prophet from the fraud.

It is clear that the situation I have described is paralleled by a stage in the actual development of religion. What men wanted at this stage was some way to find out the will of God. Individual prophets might gain a substantial following, but prophets died and their vital powers died with them. What was needed on practical grounds was a permanent record of God's will as revealed to His special prophet. And this need was eventually met by the writing of holy books, books in which God's will was revealed in a permanent fashion.

But there was more than one such book. Indeed, there were many such books. Which was to be believed? Which moral code was to be followed? Which prayers were to be recited? Which rituals were to be performed? Proofs for the existence of God are silent upon these crucial matters.

There is only one possible avenue to God's will. One must undergo a personal experience in which one senses the presence of God and apprehends which of the putative holy books is the genuine one. But it is most important not to be deceived in this experience. One must be absolutely certain that it is God whose presence one is experiencing and whose will one is apprehending. In other words, one must undergo a self-validating experience, one which carries its own guarantee of infallibility.

If one undergoes what he believes to be such an experience, he then is certain which holy book is the genuine one, and consequently he knows which actions, prayers, and rituals God wishes him to engage in. But notice that if he knows this, he has necessarily validated the existence of God, for unless he is absolutely certain that he has experienced God's presence, he cannot be sure that the message he has received is true. Thus, he has no further need for a proof of God's existence.

For one who does not undergo what he believes to be such a self-validating experience, several possibilities remain open. He may accept the validity of another person's self-validating experience. He thereby accepts the holy book which has been revealed as genuine, and he thereby also accepts the existence of God, since unless he believed that this

other person had experienced the presence of God, he would not accept this person's opinion as to which is the genuine book.

It is possible, however, that one does not accept the validity of another person's supposedly self-validating experience. This may be due either to philosophical doubts concerning the logical possibility of such an experience⁶ or simply to practical doubts that anyone has, in fact, ever undergone such an experience. In either case, adherence to a particular supernatural religion is unreasonable.

But having no adherence to a supernatural religion does not imply that one does not still face the serious moral dilemmas which are inherent in life. How are these dilemmas to be solved? To believe that God exists is of no avail, for one cannot learn His will. Therefore, one must use one's own judgment. But this need not be solely an individual effort. One may join others in a communal effort to propound and promulgate a moral code. Such a group may have its own distinctive prayers and rituals which emphasize various aspects of the group's beliefs. Such a naturalistic religious organization does not depend upon its members' belief in the existence of God, for such a belief is irrelevant to the religious aims and activities of the group.

Is it surprising then that proponents of both supernaturalist and naturalist religion are uninterested in philosophic proofs for the existence of God? Not at all. A supernaturalist believes in God because of a personal self-validating experience which has shown him (or someone he trusts) not only that God exists, but also what His will is. A philosophic proof of the existence of God is thus of no use to the supernaturalist. If the proof is shown to be valid, it merely confirms what he already knows on the much stronger evidence of personal experience. If the proof is shown to be invalid, it casts no doubt on a self-validating experience.

On the other hand, a naturalist believes either that no one has learned or that no one can learn the will of God. If, therefore, a proof for the existence of God is shown to be valid, this has no implications for the naturalist, for such a proof does not provide

him with any information which he can utilize in his religious practice. If, on the contrary, a proof for the existence of God is shown to be invalid, this casts no doubt on the naturalist's religious views, since these views have been formulated independently of a belief in the existence of God.

Who, then, is concerned with philosophic proofs for the existence of God? First, there are those who believe that if such proofs are invalid, religion is thereby undermined. This is, as I have shown, a wholly erroneous view. Neither supernaturalist nor naturalist religion depends at all upon philosophic proofs for the existence of God. To attack religion on the grounds that it cannot provide a philosophic proof for the existence of God is an instance of *ignoratio elenchi*.

Secondly, there are those who believe that if the philosophic proofs for the existence of God are invalid, our moral commitments are necessarily undermined. This is also, as I have shown, a wholly erroneous view. It is, however, a common view, and one which underlies the so-called moral argument for the existence of God. According to this argument, it is only if one believes in the existence of God that one can reasonably commit oneself to respect the importance of moral values. This argument is invalid, however, for, as I have shown, belief in the existence of God is compatible with any and all positions on moral issues. It is only if one can learn the will of God that one can derive any moral implications from His existence.

Thirdly, there are philosophers who discuss proofs for the existence of God because of the important philosophical issues which are brought to light and clarified in such discussions. So long as philosophers are aware of the purpose which their discussions serve, all is well and good. It is when philosophers and others use discussions of this sort as arguments for and against religion that they overstep their bounds. Religion may be rationally attacked or defended, but to refute philosophic proofs for the existence of God is not to attack religion, and to support philosophic proofs for the existence of God is not to defend religion.

New York University

Received October 6, 1967

⁶ Such doubts are forcefully expressed in C. B. Martin's *Religious Belief* (Ithaca, Cornell University Press, 1959), ch. V.

AMERICAN PHILOSOPHICAL QUARTERLY

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this is to include occasional collections of original articles on a common theme. These monographs can be purchased separately but are available to individual subscribers at *half price* (though not to institutional subscribers).

No. 1. STUDIES IN MORAL PHILOSOPHY. *Contents:* Kai Nielsen, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supererogation and Duties"; Lawrence Haworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, \$6.00.

No. 2. STUDIES IN LOGICAL THEORY. *Contents:* Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Honderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, \$6.00.

No. 3. STUDIES IN THE PHILOSOPHY OF SCIENCE. *Contents:* Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, \$6.00.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by

NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Alan R. Anderson
Kurt Baier
Stephen F. Barker
Monroe Beardsley
Nuel D. Belnap, Jr.
Roderick M. Chisholm
L. Jonathan Cohen
James Collins
James M. Edie
Peter Thomas Geach

Adolf Grünbaum
Carl G. Hempel
John Hospers
Raymond Klibansky
Hugues Leblanc
Ernan McMullin
Benson Mates
José Ferrater Mora
John A. Passmore

Günther Patzig
Richard H. Popkin
Richard Rorty
George A. Schrader
Michael Scriven
Wilfrid Sellars
Alexander Sesonske
Manley H. Thompson, Jr.
James F. Thomson
John W. Yolton

VOLUME 6/NUMBER 3

JULY 1969

CONTENTS

- | | | | |
|---|-----|--|-----|
| I. DAVID LYONS: <i>Rights, Claimants, and Beneficiaries</i> | 173 | VI. CHARLES B. DANIELS: <i>Personal Identity</i> | 226 |
| II. ROBERT J. RICHMAN: <i>Responsibility and the Causation of Actions</i> | 186 | VII. JAMES E. BROYLES: <i>Language and Common Sense</i> | 233 |
| III. GEORGE PITCHER: <i>Minds and Ideas in Berkeley</i> | 198 | VIII. MEIRLYS OWENS: <i>The Notion of Human Rights: A Reconsideration</i> | 240 |
| IV. NELSON PIKE: <i>Omnipotence and God's Ability to Sin</i> | 208 | IX. CHARLES LANDESMAN: <i>Actions as Universals: An Inquiry into the Metaphysics of Action</i> | 247 |
| V. JEREMY D. B. WALKER: <i>Statements and Performatives</i> | 217 | X. GEORGE DICKIE: <i>Defining Art</i> | 253 |

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$12, increasing by \$1 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$8, increasing by \$1 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4 per group of 50.

SUBSCRIPTIONS

The price *per annum* is six dollars for individual subscribers and ten dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available, and are sold at the rate of two dollars to individuals, and three dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. These monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.



I. RIGHTS, CLAIMANTS, AND BENEFICIARIES

DAVID LYONS

TO have a right, Bentham held, is to be the *beneficiary* of another's duty or obligation.¹ This theory, one of the more attractive and plausible suggestions about the nature of rights, appears supported by innumerable cases. It is reflected in the notion common to laymen, lawyers, and philosophers that someone with a right is on the advantageous side of a legal or moral relation. It promises to explain why rights are such valuable and important commodities. And it seems bolstered by a variety of facts, for example that compensation or reparation is often required, and might always be required, when one's rights are violated or infringed.

But Bentham's theory has been criticized, most notably by H. L. A. Hart, and it is not now, I think, very widely or seriously entertained. This is unfortunate, since none of the received arguments appears decisive against the beneficiary theory as such. In this paper I shall attempt to show that Hart's objections are weaker than they must at first appear because their force is largely dispelled against but one form of beneficiary theory. Another, "qualified," beneficiary theory is much less vulnerable to Hart's objections and still worthy of our consideration.

In the first section I develop and explain the beneficiary theory and distinguish two versions that might be attributed to Bentham. In the second section I argue that being a beneficiary, in a certain qualified sense, is sufficient for having a right. In the third section I chiefly discuss some difficulties surrounding the claim that being a beneficiary is a necessary condition as well.

In this paper I shall assume that straightforward statements about moral rights can be analyzed in terms of moral obligations in the same way that statements about legal rights can be analyzed in terms of legal obligations. I shall take the beneficiary theory to apply to moral as well as legal rights.²

I

Many writers have held that the notion of a right must be analyzed in terms of duty or obligation—or, more generally, in terms of requirements or prohibitions on someone's behavior.³ This is Bentham's view.

It is by imposing obligations, or by abstaining from imposing them, that rights are established or granted. . . . How can a *right* of property in land be conferred on me? It is by imposing upon everybody else the obligation of not touching its productions, &c. &c. How can I possess the *right* of going into all the streets of a city? It is because there exists no obligation which hinders me, and because everybody is bound by an obligation not to hinder me.⁴

So the idea of obligation is central to that of a right, though these ideas may be related in different ways, according to the type of right in question. Property rights, as Bentham suggests, involve others' obligations to forbear from using the property without permission. Some rights imply others' obligations even more obviously, e.g., a right to be paid ten dollars by Jones, a right to be cared for by one's parents, a right to be given equal consideration, a right not to be killed, and in such cases the state-

¹ This is my formulation, not Bentham's. It is designed to encompass the two interpretations of Bentham's theory discussed in this paper. Bentham's relevant views about rights are expressed in various ways and are spread virtually throughout his *Works* (ed. by J. Bowring, Edinburgh, 1843), which include most but not all of his published writings. I give some references below; others may be found in H. L. A. Hart, "Bentham," *British Academy Proceedings*, vol. 48 (1962), pp. 297-320. See also C. K. Ogden (ed.), *Bentham's Theory of Fictions* (London, 1932; Patterson, 1959).

Note that Bentham's theory concerns what it is to *have* a right, not what "a right" itself is; it should not be confused with theories like those of Jhering and Salmond, who hold that a right is an interest. (See *Salmond on Jurisprudence*, 11th edn., by G. Williams [London, 1957], ch. 10.)

² Bentham restricted his theory to legal rights, and in his attack on natural rights implied that there could not be any extra-legal rights. His reasons for holding the latter view seem largely independent of his analysis of rights in terms of beneficial obligations.

³ The differences between duties or obligations, on the one hand, and mere prohibitions or requirements, on the other, are touched on below. I assume in this paper that the differences between duties and obligations can be ignored.

⁴ *Works*, vol. III, p. 181; see also pp. 159ff., 217ff.

ment of the right may be held to be equivalent to the statement of the corresponding obligation. But a right to *do* (or to refrain from doing) something has as its core the *absence* of an obligation to do otherwise.⁵ Even here, however—as Bentham's second example and his general remarks suggest, and as others have agreed—one's right also seems to involve others' obligations, i.e., obligations to refrain from interfering.

Bentham held, then, that rights can be "reduced" to duties or obligations. But he did not hold the closely associated view, that rights and duties are necessarily "correlatives." That is, he did not hold that duties always imply rights. This is important to Bentham's analysis of rights. To see why, let us consider the notion of correlativity briefly.

When Bernard owes Alvin ten dollars we have equal reason to ascribe a right to Alvin (to be paid ten dollars by Bernard) and an obligation to Bernard (to pay Alvin ten dollars), and whatever would falsify one ascription would likewise falsify the other. Neither the right nor the obligation can arise without the other, and if one is discharged, waived, canceled, voided, forfeited, or otherwise extinguished the other must be extinguished as well. For the ground of the obligation—the debt—is the title of the right. Alvin's right and Bernard's obligation *necessarily coexist*, and a full statement of one logically implies a full statement of the other.

This pattern of correlations is extremely common. It obtains not only when debts are owed but also when certain other relations exist between two (or possibly more) particular individuals—as a consequence, for example, of promises and contracts, wrongful injuries that require reparation, good turns that require reciprocation, relationships such as parent to child and teacher to student. In such cases it is natural to speak of *A's* having certain rights *against B* and of *B's* having (or *owing*) corresponding obligations *to A*.⁶ And it is important that, when obligations are so grounded on such special relations or transactions and consequently can be said to be "owed" to particular persons, we can infer that the person to whom the obligation is owed has a corresponding right and that he holds it against the person with the obligation.

It is important to note what these words will be taken to signify. When *A*, in particular, holds a

certain right *against B*, *A* is a *claimant* against *B*. A "claimant" is one empowered to press or waive a claim against someone with a corresponding duty or obligation. He can, if he wishes, release the other from his obligation and cancel it, or he can insist upon its performance. A creditor, for example, is a claimant against his debtor. A promisee is a claimant against one who makes a valid and binding promise to him. So too is a person to whom a debt of reparation is owed because of wrongful injury done. A claimant is thus one to whom the performance of a duty or obligation is *owed*—he is the one who holds the claim against the other and who is entitled to administer the claim as he chooses. There are obviously moral as well as legal claimants in this sense.

The pattern of relations between rights and obligations I have just described does not seem to be universal. When behavior is simply required or prohibited by law or morals, without presupposing such special relations or transactions between particular individuals as I have mentioned, we often say that "duties" or "obligations" are imposed. But since these duties or obligations are *not* "owed" to anyone in particular, we cannot determine who, if anyone, has corresponding rights by noting to whom they are "owed." Indeed, although rights sometimes do correlate with such duties or obligations, we cannot infer *that there are* such rights merely from the fact *that there are* such duties or obligations. This point is essential to a theory like Bentham's.

Consider the following contrast. When children who have reached their majority are required by (criminal) law to support their aged and indigent parents it seems plausible to say that their parents have a legal right to such support from their children. Even though legally the parents may not be "claimants" against their children, they cannot release them from their obligations, and thus the children legally do not "owe" support *to* their parents in the full sense given above. But if children are required by law to inform the authorities of their parents' seditious remarks and activities (for which the punishment is death) we may have some hesitancy in saying that their parents have corresponding rights to be informed upon by their children. (We would be inclined to ascribe such

⁵ Or perhaps a sphere of activity in which one is free to do as one pleases. It should be observed that we are ignoring here those species of rights that lawyers call "powers" or "capacities" on the one hand and "immunities" on the other. I assume that these are rights in a different sense, or senses, of the term.

⁶ See Joel Feinberg's very helpful discussion in "Duties, Rights, and Claims," *American Philosophical Quarterly*, vol. 3 (1966), pp. 137-142.

rights only in very special circumstances.) This contrast is meant to suggest that from the fact that the law requires that *A* be treated in a certain way it does not follow, without any further assumptions, that *A* may be said to have a right to be treated in that way. That is, rights do not follow from duties or obligations, or from requirements or prohibitions, alone. Other conditions must be satisfied.

Bentham held that rights are conferred only by beneficial obligations. "To assure to individuals the possession of a certain good, is to confer a *right* upon them."⁷ Goods are assured by imposing duties or obligations, by requiring acts or forbearances of others. Obligations do not correspond to rights unless they protect or serve rather than harm or threaten those they directly concern. So Bentham held that rights "correlate" with duties or obligations in the following way, by virtue of the very notion of a right: rights imply duties, but only beneficial ones. And thus duties do not necessarily correspond to rights; they do so only when they "assure goods."

It will be useful to characterize Bentham's theory as follows: to have a right is, essentially, to be the *beneficiary* of another's duty or obligation (or of some requirement or prohibition upon another's behavior). But this needs certain qualifications before it can properly be evaluated.

Two qualifications can safely be made before considering alternative interpretations of Bentham's theory. We may note, first, that obligations are not necessarily performed. They can be breached or not fulfilled, and then the benefits or goods their performances would confer cannot be conferred in that way. Thus, a "beneficiary" in the relevant sense is not one who actually benefits but rather one who would benefit from the performance of the obligation. Secondly, a person with a right need not positively gain even when the obligation is met or discharged. For there may be indirect and irrelevant consequences resulting from (or following) the performance of an obligation. (Alvin might spend on drink the ten dollars he receives from Bernard in payment of the debt, wreck his car, and suffer serious injury—none of which would happen if Bernard fails to perform as required.) Most important, obligations may require "negative" rather than "positive" services, as Bentham says, e.g., *not* robbing, assaulting,

or killing.⁸ One may "benefit" in the sense of *not losing* rather than in the sense of gaining; what may be "assured" is not so much a good as the avoidance of an evil. The term "beneficiary" can be misleading, therefore, but we shall continue to use it with these implicit qualifications.

Beyond this point, however, Bentham's precise position is difficult to determine. Two main currents run through Bentham's discussions of rights, from which one can construct two significantly different versions of beneficiary theory.

The unqualified beneficiary theory. On the one hand Bentham seems to hold that rights are conferred whenever (and only when) rules that impose duties are justified according to the utilitarian test, that is, when the rules and the duties they impose are useful.⁹ Good laws serve the interests of individuals, and these laws confer rights. But bad laws impose duties from which no one (or hardly anyone) stands to benefit and therefore fail to confer rights. Various passages suggest that Bentham held the unqualified beneficiary theory, as attributed to him by Hart: that someone with a right is simply one "likely to benefit" or "capable of benefiting" or one who "stands to benefit" by the performance of a duty.¹⁰

This is the theory attacked by Hart. We shall examine his objections, but we need not share his alternative view in order to see that such a theory, straightforwardly understood, could not be correct. It is open to innumerable counter-examples. Suppose that Bernard owed Alvin ten dollars and also that Alvin has privately decided to give Charles a present if and only if the debt is repaid. Let us suppose further that Alvin is in no way indebted to Charles and that he has made no promises—not even tacit promises—to give Charles anything. It would seem then that Alvin has the right that corresponds to Bernard's obligation (the right to be paid ten dollars by Bernard) and that Charles has no relevant right (save the right to accept a gift from Alvin, which is clearly *independent* of Bernard's obligation). But the unqualified beneficiary theory does not differentiate between the positions of Alvin and Charles with respect to Bernard's obligation, and seems to imply that each has a right, since each "stands to benefit" from the performance of Bernard's obligation. (Indeed,

⁷ *Works*, vol. III, p. 159.

⁸ *Ibid.*, pp. 159, 181.

⁹ *Ibid.*, pp. 181, 220f.

¹⁰ See "Bentham," *op. cit.*, pp. 313–315, and also Hart's "Are There Any Natural Rights?", *The Philosophical Review*, vol. 64 (1955), pp. 180ff., where Hart criticizes the beneficiary theory without, however, attributing it to anyone.

Charles "stands to benefit" more than Alvin.) But it is absurd to suppose that Charles has a relevant right.

The qualified beneficiary theory. If the theory is to be rendered more plausible, the notion of a "beneficiary" must be refined. Two considerations seem fairly obvious. (1) Alvin's benefiting is *relevant* to Bernard's obligation in a way that Charles's benefiting is not. For Bernard's obligation directly *concerns* Alvin, that is, it requires treating *him* or behaving toward him in a certain way. But it does not similarly concern Charles, for Bernard's behavior toward Charles is not relevant to the question whether or not Bernard has met or breached his obligation. But the fact that an obligation *concerns* Alvin does not suffice to show either that he is a beneficiary in any recognizable sense or that he has a corresponding right. Brown's duty as executioner concerns the condemned Green, who (failing special circumstances) would not be called a beneficiary of the performance of Brown's duty, and who (failing certain theories of punishment) would not be said to have a corresponding right. But (2) both Alvin's benefiting and Charles's benefiting are related *contingently* to Bernard's obligation. Alvin's benefiting, or the fact that Bernard's obligation "assures" to him a certain good, *does not follow* from the obvious and literal specification of Bernard's obligation, which is simply *to pay Alvin ten dollars*. So, if the connection between Alvin's benefiting and Bernard's obligation is not *merely* accidental, and if this connection is supposed to be essential to the very notion of Alvin's right, how might it be defined?

Some remarks by Bentham suggest the view that a person with a right is a "beneficiary" in a narrower and more appropriate sense: he is *supposed* to benefit by or from the performance of another's obligation. He is the direct, intended beneficiary of that duty or obligation.¹¹

If we consider the obligations we have, and in particular those that do correspond to the rights of others, we shall find this suggestion not implausible. Viewed in an obvious and literal but, I think, somewhat superficial way, obligations do not generally require the conferring of goods *as such*; like Bernard's obligation to pay Alvin they simply require certain acts or forbearances or patterns of behavior. But we would fail to *understand* the obligations we have—we would be unable to determine

what we ought to do on the basis of them—unless we could also say whether or not those they directly concern are supposed to benefit from them. And it seems plausible to suggest that, when we fully understand the obligations we have that *do* correspond to others' *rights*, we also see that those they directly concern *are* supposed to benefit from them. This is patently true in some cases; obligations of indebtedness, reparation, and reciprocity, for example, essentially involve and require the returning or restoring of goods to particular persons. Other cases are more difficult, particularly obligations imposed by the criminal law and analogous rules of morality, on the one hand, and promissory or contractual obligations on the other. I shall deal with these separately and in some detail.

According to the qualified beneficiary theory, then, a person with a right is not one who merely "stands to benefit" from the performance of another's obligation. He is one for whom a good is "assured," or an evil obstructed, by requirements or prohibitions upon others' behavior, in the sense that some other person or persons are required to act or forbear in ways designed or intended to serve, secure, promote, or protect his interests or an interest of his. He is a beneficiary in what I shall call the "qualified" sense of the term; he is "the party to be benefited."¹² This is not to say that the conferring of certain benefits or the avoidance of specific losses or injuries is strictly *guaranteed* to him. For, aside from the fact that obligations can be breached, the specific performance of the relevant obligations might actually fail, for various reasons, to serve his interests. But I am suggesting that, on the basis of our understanding of the relation or transaction or rule that grounds the obligation, we can say in some cases that someone's interests are to be served. And, according to the qualified beneficiary theory, such persons and only such persons have rights.

II

Hart maintains, against Bentham, that:

According to the strict usage of most modern English jurists following Austin . . . the person who has a right is something more than a possible beneficiary of duty; he is the person who may, at his option, demand the execution of the duty or waive it . . . and it is neither

¹¹ See, e.g., *Works*, vol. III, p. 159.

¹² As Bentham defines "beneficiary" in his discussion of trusts: *Works*, vol. I, p. 106. But it should be emphasized that Bentham does not argue explicitly that such beneficiaries, and only such persons, have rights in the relevant sense.

necessary nor sufficient (though it is usually true) that he will also benefit from the performance of it.¹³

In this section I shall argue, with regard to Hart's criminal law examples, that the qualified beneficiary theory gives a sufficient condition for having a right (although the unqualified theory fails to do so). In the next section I extend the argument to promissory rights and counter Hart's claim that the theory fails to provide a necessary condition as well.

Hart agrees with Bentham that rights imply duties or obligations but that the latter do not always imply the former. But Hart's reasons for saying that not all duties imply rights are different from Bentham's. Bentham, as we have seen, holds that the law can and unfortunately often does impose "barren" duties or "ascetic" obligations, that is, duties or obligations without beneficiaries, which have no corresponding rights. But Hart holds that duties without claimants have no corresponding rights. And these positions are different because being a claimant is not a necessary condition for being a beneficiary of another's duty. This is shown by the fact that there are obligations with beneficiaries but without private claimants in the sense discussed above. Under the civil (as distinct from the criminal) law, private individuals are empowered to enforce their own rights; they alone can initiate proceedings against others who violate or threaten their rights. But there are no private claimants under the criminal law—only *complainants*. For the substantive rules of the criminal law impose duties or obligations that cannot be canceled by a private party; they can be "canceled" only by a change in the rules. A private individual, such as the victim of a criminal offense, can impede or aid a prosecution, but he lacks the analogous legal "power" his counterpart has under the civil law to sue. Public officials have both the legal power and the responsibility of enforcing the criminal law. Moreover, civil law obligations typically arise from special relations or transactions between the parties, whereby such obligations are incurred or assumed, but criminal law obligations are simply laid down or imposed and thus it is less natural to speak of them as being "owed" to particular individuals, and they are not "owed" in the full sense discussed above.

A similar distinction can be drawn within the

class of moral duties and obligations. Some can be incurred or assumed, by promising for example, and these can be canceled by the one to whom the promise is made and to whom its performance is owed. But some moral principles simply forbid or require, they lay down what one may not or must do, and while these may be said to impose "duties" or "obligations" they cannot always be waived or canceled.

Accordingly, Hart claims that, strictly speaking, there are no rights under the criminal law or in the analogous part of morality, because such duties or obligations are not administered by private claimants. Hart observes that jurists generally prefer not to speak of "rights" under the criminal law, for reasons akin to his. But he *also* concedes that usage differs on this point:

that a somewhat wider usage of the expression "a right" is common among non-lawyers and especially among writers on political theory who might not hesitate to say, for example, that when the *criminal* law forbids murder and assault it thereby secures to individuals a right to security of the person, even though he is in no position to waive a duty imposed by the criminal law.¹⁴

The so-called "wider" usage lacks the requirement that the person with the right (or anyone else) be a claimant. Hart goes on to suggest that such a wider use of "right" can be tolerated if it retains a central feature of the allegedly strict sense explicated by his claimant theory. We might continue to speak of rights under the criminal law provided we recognize that laws which confer rights as well as impose duties "also provide, in a distinctively distributive way, for the individual who has the right." This part of Hart's argument is somewhat difficult to interpret, but his position seems roughly the following. When the term "right" is used strictly, a person with a right has *his own* special powers as a claimant, his "limited sovereignty over the person who has the duty." This is, presumably, the "distributive" feature Hart refers to. Under a tolerable wider use, a distributive feature would be retained when talk of rights under the criminal law is confined to cases in which the law protects the security of the individual—as opposed to, say, the security of the community as a whole—that is, where a breach of such a duty "necessarily involves the infliction of harm upon a specific or (in Bentham's language) 'assignable' individual." But

¹³ "Bentham," *op. cit.*, p. 315. For criticisms of other kinds see Hart's *Definition and Theory in Jurisprudence* (Oxford, 1953). Here I deal only with Bentham's *beneficiary* theory (the first stage of Bentham's analysis of rights) and am unconcerned with, e.g., Bentham's naturalism, manifested in his further analysis of legal duty or obligation.

¹⁴ "Bentham," *op. cit.*, p. 315.

Hart claims that Bentham's analysis contains no such distributive feature, because on Bentham's theory one can qualify as having a right if he is merely "a member of a class who as a class may be indiscriminately benefited by the performance of a duty." For Bentham's theory would allegedly accord rights not only to persons protected by rules such as those forbidding murder and assault—who might be said to have "rights" under a tolerably extended usage of the term—but also to those who *might possibly* benefit by the performance of *any* useful duties, even though they do not serve the interests of individuals directly. The breach of such legal duties as income tax and military service (Hart's examples) does not necessarily involve the infliction of harm upon a specific individual, "but at the most merely makes it likely that the community as a whole will be less secure."¹⁵

Now Hart seems correct when he says that certain *ascriptions* of rights (e.g., under the criminal law) are "wider" than others (e.g., under the civil law) in the sense that the former lack implications which the latter have that the person with the right is a claimant. For rights under the civil law arise from special relations or transactions between the persons concerned and they, unlike rights under the criminal law, may be ascribed as rights "against" those having the corresponding duties, which are reciprocally "owed." (Similar remarks apply to classes of moral rights and their ascriptions.) But from this it does not follow that ascribing rights where there are no claimants involves a wider *sense of the term* "right" or uses an extended concept. For the differences between the ascriptions *as a whole* might be accounted for by the fact that some are qualified as rights "against" particular persons while others are not, depending on the sorts of conditions that warrant the ascription of the right. That is, unless we assume that being a claimant is essential to having a right in the "strict" sense of the term, we have as yet no reason to suppose that the concept of a right itself is extended when it happens not to be instantiated as a right "against" another.

Straightforward talk about rights is not generally confined to contexts in which there are private claimants. Rights may be ascribed, for example, in the context of the criminal law for the purpose of noting that some act or omission is not unlawful, e.g., when one is challenged, when interference is threatened, or to draw certain contrasts. For example, a motorist has a right to make a right

turn on a red light in California which he lacks in New York because of differences between the respective traffic laws. Such a right is not held against or with respect to any particular person, but this does not seem to make it any less of a right or a right in only an extended or loose sense of the term.

A right not to be killed may be analogous. Is it clear that we speak loosely or that we stretch our ordinary notion of a right when we ascribe, in the context of the criminal law or on the basis of general moral prohibitions, the right not to be killed? One way of finding out is by determining whether there is a single acceptable account of rights within both sorts of context, i.e., with and without private claimants. This is what the beneficiary theory purports to give.

Now let us consider more directly the force of Hart's objection against the beneficiary theory concerning rights without claimants under the criminal law. Hart maintains that rules such as those forbidding murder and assault may be said to confer rights, as the beneficiary theory implies, but only in a wider sense of the term; while rules requiring payment of income tax and military service, even when useful, cannot be said to confer rights, even in such a wider sense, although the beneficiary theory implies they do. Hart says that rights can be ascribed in the former cases because such a usage retains a certain "distributive" feature of the notion of a right, since the breach of such a duty "necessarily involves the infliction of harm upon a specific . . . individual."

But there is another and I think more obvious and plausible way of understanding why rights are and can be ascribed when murder and assault are prohibited. Such rules "assure goods" or obstruct evils to those they are supposed to protect. The qualified beneficiary theory can account for such rights without implying that every useful duty has a corresponding right. Hart's objection is directed against the unqualified beneficiary theory and, as I shall try to show, it has no force against the qualified theory.

Any rule that has a utilitarian justification and is in that sense "useful" ultimately serves (or can reasonably be expected to serve) the interest of individuals, but rules can do so in different ways. The rules that serve individuals most directly do so by imposing *beneficial* duties, in a qualified sense, which implies that there are beneficiaries in the sense employed by the qualified beneficiary theory.

¹⁵ *Ibid.*

Rules such as those forbidding murder and assault are of this type; they can only be understood as requiring that we not harm or injure others in certain ways. The duties they impose patently require treating others in ways designed or intended to serve, secure, promote, or protect their interests. The rules define the classes of persons protected, and any member of such a class is a beneficiary in the qualified sense. He does not merely "stand to benefit" by the performance of such a duty, nor does he merely "stand to suffer" if the duty is breached, for his loss at the hands of the person with the duty would be directly relevant to the question whether the duty is breached. From the point of view of the rules and the duties they impose, such a person is neither a lucky bystander in one case nor an unlucky one in the other. He is one who, according to the rules, is not to be harmed in such a way. And thus, despite the fact that such a person cannot be said to have a right "against" anyone in particular—since he cannot waive or cancel the corresponding obligation and it does not rest upon any *special* relation between him and those bound by the rules—and despite the fact that the duty is not "owed" to him in particular—since he is not a claimant—nevertheless, it does seem plausible to say that he is *entitled* to be treated in a certain way, e.g., not to be assaulted or killed, and that saying this is not speaking loosely. For others are duty-bound so to treat him, and his *right* does correspond to and correlate with their duties.

But other rules that can be justified on utilitarian grounds serve individuals much less directly. If there is a good utilitarian reason for requiring payment of income tax, for example, benefits to individuals must be expected ultimately to accrue. (As Bentham held, the interest or good of the community cannot be understood except in terms of the interests or good of its members.) But the possible, intended, or desired benefits that might ultimately accrue to individuals do not flow directly from the performance of such duties, and no harm results directly, if at all, from their breach. Duties imposed by such rules I shall call (merely) "useful." Such duties have no beneficiaries in the qualified sense. Money collected from income tax payments can be used to serve the community, and therefore its members, in various ways. But these ways are not determined by the act of payment itself. Most important, the content of the duty to pay income tax concerns payment and payment only. It does not concern the uses to which the revenue might be put. And one who might possibly benefit from the

use of such revenues, and according may "stand to benefit" from the performance of the duty to pay income tax, is not a beneficiary in the qualified sense. His benefit or loss is not directly relevant to the question whether or not the duty is discharged.

There is not even a traceable connection, normally, between someone's gain or loss as a result of payment or nonpayment and another person's discharge or breach of his duty to pay income tax. The payment marks the beginning of a long, complex chain that may, but does not necessarily, lead to benefits to individuals. Generally speaking, a particular person's payment or nonpayment is neither necessary nor sufficient for bringing about or preventing another person's ultimate benefits or losses. Usually, benefits cannot accrue in the long run unless observance of such duties is widespread. Even if it is widespread, the actual result depends on how the money is used. Suppose, however, that individuals ultimately do receive benefits that are partly traceable to income tax revenues. It remains extremely unlikely, if not impossible, that we should be able to ascribe anyone's benefits to particular performances of the duty to pay income tax. And if harm results or benefits do not accrue because payments are commonly withheld, it is again extremely unlikely, if not impossible, that particular losses could be ascribed to particular breaches. Moreover, no loss will be caused by nonpayment unless breaches are common. So it is neither necessarily the case nor even likely that anyone will suffer or lose as a consequence of one breach (or several breaches) of a merely useful duty.

Hart's objection is, then, that the unqualified beneficiary theory implies that merely useful as well as beneficial duties give rise to rights. This is because such a weak condition as "Someone stands to benefit by the performance of another's duty" *must* be satisfied whenever useful duties are imposed. But this objection has no force against the qualified beneficiary theory, since merely useful duties have no beneficiaries in the qualified sense. So, while it may be implausible to say that a merely useful duty gives rise to a corresponding right, it is also difficult to construe the qualified beneficiary theory as implying it. Hart's objection has therefore been met. Rights correspond to obligations under the criminal law in just the way Hart claims, according to the qualified beneficiary theory. Rights correlate with beneficial duties and not with merely useful ones.

Before closing this section it may be illuminating

to note that the qualified beneficiary theory accounts for rights that are related, indirectly, to such duties as military service and payment of income tax. Consider, for example, those persons who supposedly "stand to benefit" from income tax payments. More precisely, consider those who *qualify by law* as recipients of governmental services and expenditures, e.g., public education, unemployment compensation, garbage collection, and so on. The rules that govern such expenditures also provide criteria for qualification. And if one qualifies by law it would seem that he is legally *entitled* to the benefits or services and thus has a legal right to them. These rights can readily be accounted for by the qualified beneficiary theory. For *corresponding* duties—not duties to pay income tax, but duties to distribute the benefits and administer the services—fall upon those whose job it is to do such things. So, the qualified beneficiary theory seems also to account for rights that Hart does not consider. And it does so by using one purported sense of "right," which is applied in other contexts as well.¹⁶

III

Hart's more formidable objection to the beneficiary theory concerns the case of the "third-party beneficiary."¹⁷ Promises—when they are valid and binding—engender rights to promisees and obligations to promisors. A promisee often stands to benefit from the promised performance, but when a promise is meant to benefit a "third party" (i.e., one who is not a party to the agreement) then, according to Hart, the promisee as usual acquires a right even though he is not a beneficiary, while the third party, who is supposed to benefit, acquires no right. This shows that being a beneficiary is neither necessary nor sufficient for having a right.

In Hart's example, a son extracts a promise from another (let us say a friend) to care for his aged mother in his absence. Hart argues that the son has the right. The promise is made *to* him and he, therefore, has the claim against the friend. The friend's performance is *owed* to or *due* him. He alone can press or waive the claim, can insist upon its

performance or release his friend from the promise. If the promise is not kept, Hart argues, the son is *wronged* even if he is not harmed. But the mother (a third-party beneficiary) has no right. She might be harmed if the promise is not kept, but she cannot be wronged. For the promise is not made *to her*, and thus the performance is not owed to or due her. One with a right admittedly may be, and usually is, a beneficiary of a duty. But this is not what it is to have a right. The promisee alone has a moral claim upon the promisor. He is

morally in a position to determine by his choice how [the promisor] shall act and in this way to limit [the promisor's] freedom of choice; and it is this fact, not the fact that he stands to benefit, that makes it appropriate to say that he has a *right*.¹⁸

I shall try to show the limits of this objection to the qualified beneficiary theory. First, I shall argue that Hart's objection rests in part upon a misconstrual of the third-party beneficiary's position with respect to another's obligation. Because Hart considers only the unqualified beneficiary theory with its inadequate notion of a beneficiary, he assimilates the mother's position in his example to that of persons who merely "stand to benefit" by the performance of a duty and have no relevant rights. The alternative suggestion, offered by the qualified beneficiary theory, is that the mother, like those in other contexts who may properly be said to have rights although they are not claimants, can be accorded a right precisely because she is a beneficiary in the qualified sense. Then I shall examine the question whether claimants, who seem clearly to have rights, are necessarily beneficiaries—in other words, whether being a beneficiary is a necessary condition for having a right.

In Hart's discussion of the third-party beneficiary case, he concedes (as we saw he does also with regard to the criminal law) that "common usage" may sanction the ascription of rights to beneficiaries—to animals and infants for example, who are said to have rights to proper treatment because we have duties not to ill-treat them. But Hart maintains that this way of speaking employs only "the philosopher's generalized sense of 'duty'" and that it makes "an idle use of the expression 'a

¹⁶ Hart's claimant theory might be extended to accommodate rights where provision of services or benefits is conditional upon application; but it cannot accommodate cases in which there is no choice or option to exercise, e.g., where free public education is also compulsory.

¹⁷ Although Hart mentions this case in "Bentham," *op. cit.*, p. 314, his discussion of it is found in "Are There Any Natural Rights?", *op. cit.*, pp. 180ff.

¹⁸ "Are There Any Natural Rights?", *ibid.*, p. 180.

right'." He contends that "the moral situation can be simply and adequately described here by saying that it is wrong or that we ought not to ill-treat" babies or animals. But "right" and "duty" have a "specific force" in other contexts that cannot be captured by such uses of "wrong" and "ought."¹⁹ When "right" and "duty" are used strictly, and not merely in a "generalized" sense, the person with the duty may be said to be "bound to" the person with the right. The right-holder is a claimant, who controls the duty. Thus the friend has a duty "to" the son, to whom he *owes* the promised performance. But the friend has a duty "to" the mother only in the sense that his duty concerns her. The friend does not "owe" the performance to the mother, and thus she has no right to the promised services, for she has no right "against" the friend.

Hart's avowed purpose here is to draw our attention to special features of our discourse about rights, features that set it off from talk about what ought to be done or what it would be wrong to do or about "duties" that are not "owed" to claimants. He suggests that the beneficiary theory obscures the differences between these sectors of moral and legal discourse. Hart's view involves a three-fold distinction, between (a) contexts in which rights in the "strict" sense can be ascribed; (b) contexts in which rights can be ascribed only in a wider sense; and (c) contexts in which "duties" may be ascribed, but not rights. The qualified beneficiary theory does not, however, eliminate any of these distinctions (although it draws the lines somewhat differently). Corresponding to Hart's contexts (a) and (b) are those in which ascriptions of rights can be made in a single sense of the term, although the ascriptions are warranted by different sorts of conditions (depending on whether or not the duty or obligation is "owed" to the person with the right). Corresponding to Hart's context (c) is that in which duties or obligations are not beneficial and rights cannot be ascribed, even if the duties are useful. Since such distinctions between sectors of moral and legal discourse are not obscured by the qualified beneficiary theory (although they may be obscured by the unqualified theory), we need not dwell upon them.

We may consider instead what may be *common* to the various plausible ascriptions of rights to see

whether being a beneficiary is both common and essential. Hart seems to suggest here that what is common when rights are ascribed in both the strict and the wider sense of the term is that the behavior of others that is required or prohibited *concerns* the person with the right. But this condition is surely not peculiar to requirements that might be held to correspond to rights. For as we have noted, Green's duty as executioner concerns the condemned Brown, while it is doubtful that Brown acquires a corresponding right. And it concerns Brown differently than the restrictions regarding babies and animals concern them. For our duties regarding babies and animals assure goods or obstruct evils to them.

The cases of babies and animals raise complications we need not examine here, e.g., whether legal or moral personality or agency is required for the possession of rights. But it should be observed, nonetheless, that the position of the third-party beneficiary in Hart's example is different from that of babies and animals. The friend's obligation to serve the mother is not merely an instance of a general duty to refrain from harming the helpless. The friend's obligation arises from a specific agreement, and it is one he otherwise would not have. He has agreed to care for the mother. The mother is the one the promised services are intended to benefit. She is a direct, intended beneficiary.

Joel Feinberg suggests that Hart has overstated his point. Third-party beneficiaries are *sometimes* accorded rights in both morals and law. But, he says:

it does not *follow necessarily* from the fact that a person is an intended beneficiary of a promised service that he has a right to it; whereas it always follows necessarily from the fact that a person is a promisee that he has a right to what is promised.²⁰

Third-party beneficiaries can have rights, but "only in virtue of moral or judicial policies and rules." A third-party beneficiary's right seems to follow from his position as beneficiary only when, for example, it is also plausible "to say that the promisor made a *tacit promise* to the beneficiary in addition to the express promise to his promisee." The parties to the agreement let the beneficiary know of it and he thereupon "acts in reliance on its performance."²¹

¹⁹ *Ibid.*, p. 181. On related points see also Hart's "Legal and Moral Obligation" in A. I. Melden (ed.), *Essays in Moral Philosophy* (Seattle, 1958), esp. pp. 82-84, 100-105.

²⁰ "Duties, Rights, and Claims," *op. cit.*, p. 138. Assuming the premiss is valid and binding, presumably.

²¹ *Ibid.*

Here again, however, there is evidence that the objection hinges on a misconstrual of the mother's position. Feinberg's unwillingness to allow that third-party beneficiaries *necessarily* have rights seems based on a weak sense he (as well as Hart) attaches to "third-party beneficiary." When this expression is so used that it can apply even to those who, as he says, merely "stand to gain, if only indirectly" or "who will profit in merely picayune and remote ways" from the promised performance,²² a third-party beneficiary does *not* necessarily have a right. But if it is used in the qualified sense, the expression better characterizes the mother in Hart's example and it seems to *follow* that she has a right. The mother's interest is directly relevant to the friend's obligation. A complete specification of the friend's obligation includes *essential* reference to the mother, who is supposed to benefit by its performance. Her loss in case the promise is broken would not be a remote, accidental consequence of the friend's behavior; it would be the predictable and relevant consequence of a dereliction. The fact that she was not cared for would be the chief ground for saying that the friend had failed to discharge his obligation. Her receipt of the services would be the chief reason for saying that he had discharged his obligation. The friend's obligation is beneficial, and the mother is the beneficiary. The mother is *entitled* to the services required by the friend's obligation as one entitled to be treated in a certain way when another has an obligation specifically so to behave. Does she not *have a right* to be cared for by him?

Of course, in such a case the mother may not be aware of the friend's promise and thus may not know that she is entitled to his services. But knowledge of the relevant facts is surely not a necessary condition for having a right. If ignorant of the agreement, she would be in no position to complain in case the friend broke his promise. But if she became aware of the relevant facts it would seem that she could legitimately complain and "press" her right (whether or not she was *further* inconvenienced because she relied upon the friend's help). Indeed, she might also refuse the services and thus effectively "release" the friend from his obligation, especially if she could care for herself and did not need the help arranged for her by her son. But this does not mean that she has a right *because* she is a *claimant*; it means, rather, that she *may act* as a claimant (if she is in a position to do so)

precisely *because she has a right* to be cared for by the friend.

These considerations suggest once more that being a beneficiary is sufficient for having a right. But they do not show that being a beneficiary is a necessary condition, and thus the more serious threat posed by Hart's objection has not been met. For the son, as promisee, would seem to have a right against the friend. But the son does not appear to be a *beneficiary* in any straightforward sense. The promise is meant to benefit the mother, not the son, and she alone will benefit directly from the friend's care. Moreover, it is the friend's treatment of the mother, not his treatment of the son, that is relevant to the obligation, that determines whether or not it is discharged. This seems to indicate that being a beneficiary of a beneficial obligation is not a necessary condition of having a right.

If this part of Hart's objection is allowed, a defender of the beneficiary theory might be content with the following observations. The argument has tended to show that being a beneficiary of a beneficial obligation is, first of all, a sufficient condition for having a right. Hart's objection shows at best that being a beneficiary is not a necessary condition for having a right. But it does not show that rights are conferred even when there are no such beneficiaries. And this may not be an insignificant point. For it would seem that outside the contexts of promises (contracts, agreements) counter-examples to the qualified beneficiary theory are not easy to find. In other contexts in both law and morals it would seem that rights are ascribed when and only when obligations or restrictions upon others' behavior assure goods or obstruct evils to individuals and thus when breaches of the corresponding obligations would involve loss or harm to those with the rights. The apparent possibility of counter-examples arises because it does not seem to be a condition of a valid and binding promise that its performance serve the interests of the *promisee*. But in such cases there is at least a *third-party* beneficiary. It is difficult to imagine promises—or, generally, obligations that correspond to rights—which are valid and binding and yet serve no one's interests. On the contrary, our understanding of the nature of binding promises is shown by the fact that, when their fulfillment threatens unexpected disadvantages to those they are meant to serve, it would be wrong to keep them.

²² *Ibid.*, p. 137; see also p. 138, including note 3.

But if we pursue this line of argument and attempt to clarify the respects in which promises may be said to "serve someone's interests," then it would also appear that a promisee, even in a third-party beneficiary case, is a "beneficiary" in the qualified sense.

Suppose that Jones extracts a promise from Brown. The promise is not intended to benefit a third party, and there is no one who could plausibly be regarded as a "third-party beneficiary" in the appropriate sense. Brown makes the promise because he is led by Jones to believe that he, Jones, wants the promised act performed. But suppose that Jones is, in fact, unconcerned whether or not the promise is kept. He extracts the promise on the merest whim or out of malice. It seems clear that the conditions of a valid and binding promise are not satisfied (whether or not Brown knows this), and thus that Jones does not acquire a *right* to what was promised and that Brown incurs no *obligation* (although he may *think* that he is obligated). But if Jones had really wanted what was promised and Brown had freely agreed to do it, then—barring immoral purposes—a valid and binding promise would have been made.

What does this show? One of the conditions of a valid and binding promise, and thus a condition of a right accruing to the promisee, is that he really wants what is promised.²³ Now this does not imply that there must be *benefits* in any straightforward sense. But it does, I think, imply that the obligation "assures a good" to the promisee. For it is the promisee's very want, wish, or desire *to have what is promised done* that the promise is *meant* to satisfy. And it is not implausible to suggest that the satisfaction of someone's (morally permissible) want, wish, or desire amounts to the conferring on him of a certain good.

If we try to imagine cases in which this claim is not satisfied, we shall find the ascription of a right and of a corresponding promissory obligation moot. Suppose, for example, that the promisee asks another to do him an injury or to kill him. We can imagine cases where, to avoid greater evil or unbearable pain, for example, such a promise might be reasonably requested and made and consequently binding. But in such cases there is also as much reason to say that the promise "assures a certain good" to the promisee. If there is no such good reason for the promise but only, say, a desire for self-destruction or mutilation, it is not at all

clear that one would be morally free to make such a promise or that once made the promise could be regarded as morally binding.

Let us return now to Hart's example. If a son extracts a promise from his friend to care for his aged mother in his absence, we might reasonably suppose that he does so because, say, he is concerned, as an affectionate son, about his mother's well-being; or, perhaps, because he thinks he is serving his own interests in providing for her care; or even because he wants to discharge what is to him a totally disagreeable filial obligation. That is, we might reasonably suppose that, for one reason or another, the son really wants his mother cared for in his absence, that he has *some* interest in her being cared for, and that he extracts the promise accordingly. But let us not assume any such thing. Let us suppose instead that the son merely wants his friend to make such a promise, but that it is not part of his intention to provide for her care and that he has no interest in doing so.

Now, on the one hand, it might transpire that the friend acquires an obligation to care for the mother and that the mother acquires a right to the promised services—that the promise is binding to this degree, despite the son's secret indifference. For the mother has an interest that needs to be served and the friend freely agrees to serve it. But on the other hand, it would not seem that the son is truly entitled to complain if the friend fails to keep his promise. And that is because he has not satisfied the conditions necessary for acquiring such a right. If there is an obligation, it is not owed to the son (whether or not the friend knows it). If there is a legitimate claim it is not the son's claim but the mother's; and if there is a right it is not his either.

If there is such a difference between the son's moral status in this case and in Hart's original example, so that it is more plausible to ascribe a right to him in the latter than in the former, that difference has to do with the son's reasons for and his sincerity in extracting the promise. What is true of Hart's example that is not true here? The son wants his mother cared for, and for that reason asks his friend to do so. The reason he gives his friend is the same in both cases; but only in Hart's example is that really his reason. So, in both cases at least part of the point of the promise is to satisfy such a desire or wish on the part of the son—to provide for his mother's care—but only in Hart's

²³ For a similar point concerning promissory obligations, see Jerome Schneewind, "A Note on Promising," *Philosophical Studies*, vol. 17 (1966), pp. 33-35.

case will the keeping of the promise actually satisfy such a wish or desire.

This is not to say that the difference between the two cases, which may account for the fact that the son acquires a right in one but not the other, is the fact that in one but not the other the son "stands to benefit" from the promised performance. Nevertheless, the fact that the son has an interest in his mother's care and wants her cared for, coupled with the fact that it is *this* wish or desire of the son's that the friend's promise is designed to satisfy, shows that the friend's obligation "assures a good" to the son.

I do not mean to claim that this defense of the beneficiary theory is conclusive. I hope to have shown, rather, that the theory cannot be dismissed without further consideration. It is less vulnerable than its critics have supposed, largely because of the differences between its qualified and unqualified versions.

I shall conclude by considering an example that may indicate serious difficulties for the beneficiary theory. The example is suggested (in another connection) by Hart, who maintains that a certain class of moral rights correlates with "political obligations"—obligations to conform to certain social rules. These rights and obligations arise, Hart says, in the following circumstances:

when a number of persons conduct any joint enterprise according to rules and thus restrict their liberty, those who have submitted to these restrictions when required have a right to a similar submission from those who have benefited by their submission. The rules may provide that officials should have authority to enforce obedience and make further rules, and this will create a structure of legal rights and duties, but the moral obligation to obey the rules in such circumstances is due to the co-operating members of the society, and they have the correlative moral rights to obedience.²⁴

As I have argued elsewhere,²⁵ certain qualifications should be added, at least to make the character of the present argument clear. These include (a) actual conformity to the rules is sufficiently widespread to produce some shareable good or to prevent some common evil; (b) this desirable result could not be achieved without such cooperation; (c) the benefits and burdens are fairly distributed; (d) the total benefits outweigh the burdens imposed; (e) *universal* cooperation by those who stand to benefit is *not* required to achieve the desirable end. The last condition is most important, because

it explains how it is possible for a "freeloader" to take advantage of others' cooperation without himself performing as the rules require. The conditions also show that the duties imposed by the rules and the moral obligations to conform to them are not beneficial in the qualified sense employed above. For a single breach of such a rule is not sufficient for anyone's loss, and the performances required by these duties and obligations do not consist in directly serving the interests of someone with a correlative right.

The sort of social rule that might satisfy such conditions as I have outlined is the legal rule requiring payment of income tax. This rule imposes a legal duty to pay income tax and, if the conditions are satisfied, then, according to the argument given, one has a moral obligation to conform to such a rule and others have a right to one's conformity or cooperation. Consequently, the qualified beneficiary theory cannot account for such rights, since the legal duties in question are merely useful and the same holds for the moral obligation to obey such a law.

What are the connections between goods and rights in this case? On the one hand, there cannot be such an obligation unless the law, or system of law, is useful (as conditions (a), (b), and (d) require), and unless everyone stands to benefit (as conditions (c) would I think require). So goods *are* essentially involved; but they are involved indirectly, as in the case of merely useful duties. For (as condition (e) seems to guarantee), the obligation to obey the law is *not* contingent on universal obedience being needed for the usefulness of the general practice. The idea that underlies such an obligation and its correlative right is that it would be *unfair* or *unjust* for someone who benefits from others' burdensome cooperation or conformity to fail to perform when his turn came. The idea is not that his failure to conform or co-operate would detract from the usefulness of the general practice or would decrease in any way the benefits that accrue to individuals. For according to the conditions of the argument, even, say, the secret and harmless breaking of necessary rationing restrictions would be wrong and would involve the violation of such a right and the breach of such an obligation.

If this is correct, then it would seem that the only good that can be assured by such an obligation to obey the law in *each* of its instances is the *abstract*

²⁴ "Are There Any Natural Rights?", *op. cit.*, p. 185.

²⁵ *Forms and Limits of Utilitarianism* (Oxford, 1965), ch. 5, Sect. A.

or *impersonal* good of justice. It is not a personal good, not a good *to* or *for* a person with a corresponding right. And this does not satisfy the conditions of the qualified beneficiary theory.

Utilitarians would, of course, deny that there is such an obligation, or would insist that it rests on the fact that nonconformity or noncooperation can reasonably be expected to have bad results, by

setting a bad example which may lead others to disobey, thus tending eventually to decrease the usefulness of the law. This suggests that the good assured to individuals is simply the confidence that their burdensome cooperation will not be in vain. Alternatively, utilitarians may seek to argue that the abstract or impersonal good of justice reduces, ultimately, to personal goods.²⁶

Cornell University

Received September 11, 1968

²⁶ Earlier versions of this paper were read at Cornell University and the Universities of Waterloo, Massachusetts, Pittsburgh, Michigan, and Texas, where I received helpful comments and criticisms. I have also benefited greatly from discussions with Norman Malcolm and Michael Stocker and from the suggestions of Ellen Coleman, Robert Monk, William Nelson, and John Turner.

II. RESPONSIBILITY AND THE CAUSATION OF ACTIONS

ROBERT J. RICHMAN

THIS paper is divided into three sections. In the first, I briefly set forth a familiar and fundamental philosophical puzzle which in the paper is labeled the *Dilemma of Responsibility*. In the second, I argue that contemporary analyses of the concept of human action which claim that the category of causality is inapplicable to these actions is inadequate for a resolution of the Dilemma. Finally, in the third section I attempt to provide a resolution of the Dilemma which does not require the denial of the traditional supposition that the category of causation may sensibly be applied to human actions. Regarding the causation of actions, I should make clear at the outset that I am not primarily concerned with the question of whether desires, say, are (really) causes of actions. As with most philosophical questions of this form, there are considerations to be adduced on either side, none of which is conclusive. In the second section of this paper I shall argue that denying that desires are causes of actions does not help to resolve the Dilemma of Responsibility; in the third, I shall argue that affirming that desires are causes of actions—as I shall do for the sake of the argument—does not prevent a resolution of the Dilemma.

I

The Dilemma of Responsibility arises out of a quarrel between determinists, roughly, those who hold that every event—or, at least, every macroscopic event—has a cause, and indeterminists, who claim that some events (and, in particular, some human actions) are uncaused. Both determinists and indeterminists claim that their opponents' position is incompatible with human responsibility. A combination of these claims yields the Dilemma of Responsibility. The horns of this dilemma are boringly familiar; I shall set them forth simply and rather uncritically.

(1). The indeterminist claim: If every event, including human actions, has a cause, then we cannot make sense of the notion of responsibility. For,

given the antecedent causal conditions of an action the agent *must* (causally) perform it. But a person can sensibly be held responsible for an action only if it is causally possible for him either to perform or to refrain from performing it. (Cf: "*Ought* implies *can*.")

(2). The determinist claim: If (some) human actions are uncaused, then we cannot make sense of the notion of responsibility as applied to these actions. For, as it is sometimes put, if the agent—or his "character"—is not the cause of a given action, then the action is not *his*. Or, to make the point a bit differently, if an action does not result from an agent's desires, etc., then the "action" is a mere happening, and not a genuine action; in any event it is not attributable to the agent.

I shall not here discuss the standard counter-moves to these claims, although I shall refer to several of them below.* Many of these moves represent important aspects of our actual moral thinking. None, however, seems to me effectively to undermine the metaphysically oriented view against which it is raised. This should not surprise us when we consider that the concept of responsibility arises and functions in the context of the social control of behavior, not in that of the resolution of theoretical puzzles. I shall return to this point in the third section of this paper. Now I should like to examine the efficacy of contemporary action analysis in resolving the Dilemma of Responsibility.

II

Discussion of the concept of human action has been a prominent aspect of recent Anglo-American philosophy. This discussion has yielded many valuable insights, but I am concerned with it only as it bears on the resolution of our problems. For brevity, I shall use the term "action analysis" for any analysis which distinguishes between human actions and bodily movements, between, e.g., one's raising one's arm and one's arm rising, and which

maintains that causal explanations are applicable to the latter but not to the former. Proponents of action analysis, of whom Professor A. I. Melden in *Free Action* (London, 1961) may be taken as representative, will be called *action analysts*.

Action analysts are generally at pains to argue for the inadequacy of one traditional view which equates actions with bodily movements plus some mental accompaniments, e.g., "volitions" or desires. The inadequacy of such an analysis does not imply that the agent's desires—"volitions" may well be a philosopher's myth—are irrelevant to the explanation of his actions. Indeed, for "full-blown" cases of actions, especially for the most interesting and important cases of responsible actions, reference to the agent's desires appears essential for purposes of explanation. This point is not denied by action analysts; only, they claim, explanations of actions are never *causal* explanations, and, more particularly, the agent's desire to do something is never (cannot be) the *cause* of his acting.

If we take the Dilemma of Responsibility seriously, it seems to me that the two points of action analysis which I have singled out, viz., the sharp distinction between actions and bodily movements, and the claim that desires cannot be causes of actions render this analysis incapable of resolving our Dilemma.

Before turning to these two points, however, we should be a bit clearer about the relevance of action analysis to the Dilemma. We should be aware, for one thing, that the action analyst's denial that causal explanations of human actions are possible is not to be equated with ordinary indeterminist views. According to these views, many, if not most, human actions are causally explicable in terms of the agent's desires. Even if the indeterminist were to claim that no actions have causes, this claim would be one of fact, and not of conceptual necessity, and in this respect it would differ from that of the action analyst. In any event, the merely negative claim that (some) actions are uncaused leaves them a complete mystery, and, moreover, strongly suggests the radically random character of (these) actions which the determinist sees as the sole alternative to his account. In other words, such a claim provides no account of actions, and is of no account in explaining responsibility for actions. An attempt at a more positive indeterminist—or "libertarian"—account, such as that of C. A. Campbell, grants

that the vast majority of human actions are causally determined, but claims that free will operates in a limited sector in which one's strongest desire clashes with one's sense of duty. According to Campbell, "... it is precisely that small sector of the field of choices which our principle of delimitation leaves open to free will—the sector in which strongest desire clashes with duty—that is crucial for moral responsibility."¹ Campbell's account leaves his central case in which an agent "rises to his duty" a complete mystery. Since, for Campbell, determinism is incompatible with responsibility, and since, moreover, he accounts desires as causes of actions, the obvious explanation that an agent may desire to do what he conceives to be his duty is ruled out. Apparently, adequate moral training would render an agent incapable of performing a praiseworthy action, since such training would *incline* one to, say, benevolence. Worse, if an agent performed a heinous act out of desire or inclination, and without any consideration of his duty, the action would not fall in the sector which is "crucial for moral responsibility." Clearly, Campbell's account too narrowly delimits the area of moral responsibility. It is, moreover, inadequate as an account of the sector of human action to which it applies.

To repeat, action analysis is not to be identified with indeterminism. Nor, clearly, is it a form of—or indeed compatible with—determinism. The action analyst deals with the Dilemma of Responsibility not by accepting either horn, but by maintaining that the question underlying both determinism and indeterminism, namely, "Do human actions have causes?", is radically misguided. On his view, accepting this question as significant leads us into insoluble perplexity.

For our present problem, then, the fundamental claim of the action analyst is that causes cannot significantly be assigned to actions. On this analysis, factors which function in the explanation of actions, e.g., the agent's desires and beliefs, are not causal factors. For purposes of simplifying the discussion, I shall, throughout this paper, refer to the agent's *desires* as the significant factor in the explanation of his behavior. Other operative factors, notably the agent's beliefs, may be thought of schematically as part of the standing conditions or circumstances in which the action occurs.

The action analyst has an argument for his claim, that desires are not causes of actions. Although, as will appear, I find the argument far from per-

¹ "Is 'Freewill' a Pseudo-Problem?", *Mind*, vol. 60 (1951), p. 461.

suasive, I should make clear that I am not so much concerned with the cogency of the argument as with the adequacy of its conclusion for dealing with the Dilemma of Responsibility.

It is claimed that a desire cannot be the cause of an action because a cause and its effect must be logically independent, while desires and actions are logically related. The doctrine that a cause must be logically distinct from its effect is attributed to Hume.² One may well wonder whether Hume conceived as *logical* any relation other than entailment and concepts interdefinable with it. (And surely no one maintains that the occurrence of a desire entails the occurrence of the action desired.) Even if Hume conceived of a weaker logical relation than entailment, moreover, unless we accept Hume's account of causation as definitive, we may hesitate to accept the doctrine in question. We can, of course, stipulate that nothing shall count as a cause which is not—in a sense not yet specified—"logically independent" of its effect, but then the question will arise as to whether all standard cases of causes satisfy this stipulation.

The action analyst's argument proceeds with the claim (stipulation?) that logically independent events must be capable of independent description, i.e., that if two events are logically independent, then it must be possible to give a specification of one which does not involve a reference to the other. Now *prima facie* it would appear that the logical independence of events involves only the possibility of their independent occurrence—a feature which clearly characterizes desires to act and actions. That logically independent events must be independently describable may, however, be construed as part of the action analyst's conception of *logical independence*. But if we accept this construction, then the claim that causes and effect *must be* logically independent seems hardly a self-evident truth.

Finally, according to the argument in question, a desire to act and the resulting action cannot be logically independent, since a description of the desire necessarily makes reference to the action desired. This last point is not obviously true. It would appear that some desires could be singled out by reference to the time of their inception. More speculatively—and more interestingly—the possibility cannot be ruled out *a priori* of the establishment of correlations between certain desires and certain brain states. Given this finding, some desires could be described by reference to

their related brain states. To the anticipated objection that desires so described fail to *explain* actions, the response must be given that in the requisite sense Humean causal statements based solely on contiguity and succession of events generally fail to provide *explanations*.

In any event, "*X* did *A* because *X* wanted to do *A*" hardly provides the model schema for the explanation of an action in terms of a desire. When we consider actual cases of explanations of this sort, we find that while it is true that the desires are specified by reference to their objects, the objects in question are rarely just the actions to be explained. More commonly they are states of affairs to which the actions to be explained stand in a means-ends relation, or alternative descriptions of the actions in question. Hence, if it is claimed that the explanation of an action in terms of a desire requires that the specification of the desire makes a reference to the action, this reference is at most indirect, and is required only on some presupposed model of "complete explanation." When I say that I am swallowing aspirin because I want to be rid of a headache, the description of the desire makes no direct reference to the action.

Another sort of example will perhaps show that the demand for independent describability of cause and effect—even in the "Humean sense"—is of doubtful validity. Suppose I discover that whenever I utter a statement of the form "I wish that *X* were dead," *X* dies immediately afterwards ("of natural causes," we may add). Experimenting in a detached manner, I find that my use of the utterance in question is invariably "successful." Surely, after a time I should be justified in concluding that (in some mysterious way) my utterances were causing the deaths of the unfortunate persons involved. I should have to become very careful—or at least very selective—in what I said. But in the action analyst's sense, the utterances and the deaths would not be logically independent because they are not independently describable. It would seem, therefore, that the logical requirement of independent describability of causally related events is unwarranted.

Although the argument designed to show that desires to act and actions cannot be causally related fails, it may nonetheless be true that, as the action analyst claims, desires and actions are conceptually related. Surely, the relation between desires and actions seems to be stronger than that of an empirically ascertained statistical correlation,

² See e.g., Melden, *Free Action*, *op. cit.*, pp. 53 and 105.

although, of course, weaker than that of entailment. One plausible way of specifying the sort of relation in question is to say that there is a *conceptual relation* holding between *X* and *Y* if and only if in the absence of specifiable conditions to the contrary, the existence or the occurrence of *Y* may justifiably be inferred from that of *X*. Unfortunately, this specification fails to distinguish conceptual from causal relations. For causal generalizations provide grounds for inference and, moreover, they must generally be fitted out with *ceteris paribus* clauses. If the claim that *X* and *Y* are conceptually related is to have force, it must imply that *ceteris paribus*, i.e., in the absence of specifiable countervailing reasons, *X* is necessarily, or in virtue of the concepts involved, accompanied by *Y*. But on this reading, the claims that desires and actions are conceptually related is of no help in resolving the Dilemma of Responsibility. It will be recalled that the difficulty with the view that desires and actions are causally related is that, given the occurrence of the desire to perform a certain action, it becomes—*ceteris paribus*—causally impossible for the agent to refrain from performing the action. The analogous difficulty with the view that desires and actions are conceptually related is that, given the occurrence of the desire to perform a given action, it is—*ceteris paribus*—conceptually impossible for the agent to refrain from performing the action. This is far from a solution of the problem.

It may be claimed that the proposed interpretation of the notion of *conceptual relation* does not accord with the intention of the action analyst. Clarification of that intention would be welcome. Meanwhile, the difficulty raised in the preceding paragraph may be formulated in more familiar terms. It is by no means implausible to make the (conceptual) claim that, given an agent's ability to perform or to refrain from performing a given action, together with his present desire to perform the action, the only possible obstacle to his performing it is some stronger contrary desire. But one's desires and abilities—at least at a given moment of time—are not a matter of one's choice; they are not "up to" an agent in the way in which his actions are presumed to be. To be sure, had the agent's earlier actions been different, his present desires and abilities would have been—or, at least, might have been—different, but the difficulties which apply to his present actions apply equally to his earlier ones. It appears that the claim of there being a conceptual relation between

desires and actions raises the difficulties of the Dilemma of Responsibility in a new form. If, however, this conceptual relation is denied, action analysis leaves us with the Dilemma in its original form.

Even, however, if action analysis provided a completely satisfactory account of the relation between desires and actions, it would still face another difficulty, this one based on its distinction of actions from bodily movements. For the sake of argument, we may ignore the question of the causation of actions, or even suppose that it is devoid of significance. Still, when we raise the question of the causation of bodily movements, the Dilemma of Responsibility again arises to haunt us.

Bodily movements may be characterized, at least roughly, as motions of a (human) body, or of any of its parts, e.g., the change of position of John's body from one side of the room to the other, Henry's arm going up, Marvin's lips parting, or Frank's vocal chords vibrating. Clearly, most human actions involve bodily movements: John can't cross the room without his body's moving from one side of the room to the other, and Henry can't raise his arm without his arm's rising. The qualification "most" is designed to deal with those cases which we account actions, e.g., refraining from voting, keeping silent, or standing motionless, which do not involve specific bodily movements (although they rule some out). It will simplify our discussion, without distorting it, if we omit consideration of such cases, and consider only those actions which essentially involve bodily movements.

The relation between actions and bodily movements is of central importance for our problem. It seems clear that, as the action analyst insists, actions are not to be characterized as bodily movements plus mental accompaniments. On this point it is perhaps sufficient to note that we can normally tell what an agent is doing (signing his name, playing a Beethoven sonata, opening a can of sardines) without having to inquire into his accompanying mental state. Difficulties in the view that actions are to be analyzed as bodily movements plus mental accompaniments are familiar,³ and I shall not review them here.

Rejection of this view is thought to be of importance for action analysis, since if every action necessarily involved independently identifiable mental events, it might be claimed that these mental events are causes. But, one may wonder,

³ *Ibid.*, esp ch. 8.

causes of what? On the view rejected, these events would be components of actions and, hence, not their causes—at least not on the action analyst's use of "cause." But the action analyst wants to maintain only that *actions* cannot significantly be said to have causes: he is more than willing to allow that the question of the causation of bodily movements is significant.

But if bodily movements have causes—or if they do not—then, given the relationship between these movements and actions, namely, that under suitable circumstances a bodily movement *is* an action, we are confronted again with the Dilemma of Responsibility. Elimination from the analysis of the concept of action of reference to accompanying mental states removes one source of obfuscation. But one thing which this elimination makes clearer is the difficulty which confronts us. Consider any bodily movement *M*. Either there are sufficient causal conditions for the occurrence of *M*, or there are not. If there are, then given the causal conditions, *M* cannot but occur.⁴ If there are not, then *M* is a chance or random event. In neither case can *M* be reasonably claimed to be under the control of an agent. Suppose now that *M* occurs under circumstances such that *M* constitutes an action. Since *M* is either a random event or else the causal consequence of events over which no agent has present control, it is not reasonable to hold the individual whose bodily movement *M* is, accountable for *M*, or consequently for the action which is constituted by *M*.⁵

The foregoing dilemma is set forth solely to argue for the point that if there is a genuine Dilemma of Responsibility, then it is not to be resolved by the action analyst's distinction of actions and bodily movements, with the category of causation being applicable only to members of the latter class. It is *not* employed to show that human beings must be—deterministic or non-deterministic—automata. Nor, surely, is it used to argue against the obvious fact that human beings are capable of initiating actions. Anyone needing reassurance on this point need merely start to twiddle his thumbs (if he is not already doing so).

As a postscript to this discussion of action analysis, I should like to agree with the action analyst's claim that, in general, desires are not independently identifiable events. This agreement,

however, is based not on conceptual grounds but on experience. Normally, one's desires are manifested in one's behavior; they do not occur as events distinct from it. This is so largely because most of our behavior does not involve a great deal of ratiocination. For the most part our behavior proceeds smoothly enough without the need for the conscious awareness of ends or for the calculation of means to these ends. The philosopher is normally concerned with rational actions, and his account of *actions* tends to be a "rational reconstruction" which represents an idealized case of rational actions. Clearly this fact is of no comfort to the action analyst or to anyone else who is attempting to deny the *caused* character of human actions. Surely the bypassing of conscious desires and beliefs in the case of most actions is no basis for asserting that these actions are uncaused. Indeed, it is just the distinction of reasons for actions, especially desires and beliefs, from other, more obviously causal, factors which is essential to the action analyst's case. I shall proceed, therefore, to talk a good deal about beliefs and, especially, about desires, recognizing the idealized and schematic nature of the resulting discussion. It may be helpful for this discussion to think of a desire as represented by any true (not necessarily truthful) answer which might be given to a question of the form "Why did you do—?" and which is embodied in a statement of the form "Because I wanted—." Neither the agent's conscious thinking of this answer nor his recognition of its truth when it is thus considered is essential to this conception of desires. Thus, without commitment to any psychological theory, this conception allows two levels (or senses) of "unconscious desires."

III

How are we to resolve the Dilemma of Responsibility? At a purely abstract level, I should urge, the Dilemma may admit of no resolution. Insofar as this is the case, however, two points should be noted. First, at this highly abstract level the Dilemma employs the concept of responsibility in such a way that it can have no *possible* application. Hence, the Dilemma is not directly relevant to our *practical* concern with responsibility. Second, so long as the Dilemma remains unresolved, it is

⁴ For a more detailed account of the situation here envisaged, see Carl Ginet's "Might We Have No Choice?" in *Freedom and Determinism*, ed. by Keith Lehrer (New York, 1966), pp. 88–92.

⁵ I have developed this argument more fully in "Reasons and Causes: Some Puzzles," forthcoming in the *Australasian Journal of Philosophy*.

illegitimate to seek to impale one's philosophical opponents on one of its horns. Thus, e.g., some proponents of the Freewill Defense⁶ argue that God is to be absolved of the responsibility for evil in the world because this evil may result from the uncaused actions of humans (and other—even more mysterious—beings). The argument rests on the claim that in a completely deterministic world, beings—even if they satisfied the criteria of freedom and responsibility which we normally employ—would really be automata, lacking genuine freedom. This claim may be recognized as roughly equivalent to one half of our Dilemma. But the argument is without force until its proponents explain how they deal with the other half. Thus, if beings just like us, satisfying the common-sense tests of responsibility, might nonetheless be automata (or “quasi-automata”), then their possession of contra-causal “freedom” would not alter their status. An automaton some of whose movements are randomized remains an automaton, and becomes thereby in no interesting sense “free” or “responsible.”

If we bear in mind that the primary use of the concept of responsibility is a practical one, we should not be too surprised to find that the concept is subjected to great, and perhaps unbearable, stress when it is removed to the abstract metaphysical context of the determinism-indeterminism controversy. Clearly, the truth or falsity of determinism forms no part of our everyday criterion of responsibility: in attempting to decide whether an individual is responsible for a given action we do not, as part of our normal deliberation, try to answer the abstract question of whether or not every event has a cause. Still, it may be said, the truth of indeterminism (or of determinism) is—if not a criterion of responsibility—at least a necessary condition of the existence of responsibility. It is a necessary condition, the claim runs, because the denial of the doctrine is incompatible with the existence of responsibility. If, however, determinism and indeterminism are so defined as to be contradictory views, and if proponents of each view produce cogent arguments for the incompatibility of their opponent's position with the existence of responsibility, then we must conclude either that determinism and indeterminism are incoherent doctrines, or else that the notion of responsibility appealed to in the arguments is logically incapable of application. Although determinism and indeterminism are by no means clear doctrines, and

are almost certainly—taken singly, of course—unverifiable, still, it appears, they are not totally incoherent. They are clear enough, at least, so that certain inferences can be drawn from them. I incline strongly, therefore, to the second alternative, that there is a notion of responsibility—an extension of our ordinary practical notion, which may be called *ultimate* or *metaphysical* responsibility, and which has no possible application.

But if the soundness of the Dilemma of Responsibility depends on a metaphysical or a “strict” sense of “responsibility” (“In a strict sense no one is ever responsible for his actions”; cf., “In a strict sense no one can ever *know* any matter of empirical fact”), still this fact does not show that there is no sense (a “loose” sense?) in which the term is applicable. After all, we do manage to use the term successfully. But here we must beware of providing too simple a “solution” to our philosophical problem, that of claiming that our difficulty arises from a mere equivocation, from the employment of two different senses of “responsible” (or of “know” in the case mentioned above). This move fails to explain the perennially felt perplexity of the philosophical problem. The notion of differences (or sameness) of sense or of meaning is much too unclear generally to account for and to resolve serious philosophical perplexity. Are we to say that in the epistemological case indicated, the sceptic employs a different concept of knowledge from the ordinary, or that he uses the ordinary concept in an extended way, namely, by placing abnormally strict conditions on its application? The two formulations may come to much the same thing, but the latter at least has the merit of indicating a relationship between the two “senses” of “knowledge.” If this relationship is not recognized, then the philosophical puzzle appears to be merely silly.

Similarly with our problem. It won't do to say simply that the truth or falsity of determinism forms no part of the ordinary criterion of responsibility, and that therefore a different *meaning* of “responsibility” is employed in the Dilemma from that which is employed in everyday life. Unless “criterion” is taken as meaning something like *necessary* or *defining characteristic*, it is by no means obvious that difference of criterion entails difference of meaning. Indeed, in its more usual sense of (fundamental) test-condition, it seems obvious that the criterion for the application of a term may (and not infrequently does) change, without a concomitant change in the meaning of the term.

⁶ E.g., Alvin Plantinga, “The Freewill Defense” in *Philosophy in America*, ed. by Max Black (Ithaca, N.Y., 1965), pp. 204–220.

(Despite current philosophic preoccupation with the "use" of terms—a preoccupation which may obscure this and other distinctions—one may say that the adequacy of criteria for the application of a term depends, in part, on the term's meaning. Whereas it makes sense to question the adequacy of criteria, it does not in the same way make sense to question the adequacy of the meanings of terms.) I use the term "criterion" to refer to a fundamental test-condition, but this decision is not of fundamental importance. One could use the term to refer to an essential or logically necessary characteristic. And on this usage, of course, a difference in criteria would imply a difference in meaning. Only, we must not confuse the two uses of "criterion." In particular, we must not suppose that a difference in test-conditions (i.e., a difference in *criteria* in the first sense) implies a difference in meaning (*via* a difference in *criteria* in the second sense).

Rejecting as not wholly adequate the "two senses" solution to our problem, I should like to look at the ordinary criteria (i.e., test conditions) for the attribution of responsibility for actions, and at the manner in which consideration of these criteria leads us into our philosophical Dilemma. I shall be concerned to indicate only some of the major points of this development. First, I should make clear that to attribute to an agent responsibility for an action is not—in the sense with which I am concerned—*merely* to say that he performed the action. It is to say, in addition, that he is accountable for the action, and that, at least in the more interesting cases, he is deserving of praise or blame, and perhaps even of reward or punishment, for its performance. I should maintain that if by criteria of responsibility are meant some general set of characteristics or circumstances of actions, then in this sense there are no specifiable criteria of responsibility for actions. Rather, in terms of our ordinary criteria, an agent is responsible for each of his actions unless there are circumstances attending the action which absolve him of responsibility. Such circumstances may, for the sake of brevity, be called *absolving* conditions. In the case of an action which is *prima facie* blameworthy, absolving conditions would naturally be called *excusing* conditions. Excusing conditions are to be distinguished from *extenuating circumstances*, in that the latter merely reduce the degree of blameworthiness for an action without eliminating responsibility for its performance.

Absolving conditions are one of a philosophically

important set of what may be called *defeating conditions*. A description of a defeating condition may be thought of as replacing the dash in statements of the form "A ψ is a ϕ unless—." The notion of defeating conditions, or of sufficient reasons to the contrary, is obviously of great importance in the analysis of moral judgments, e.g., "Killing is wrong unless—" or "Kindness is good unless—." It also has relevance to many other areas of philosophic inquiry—to any, indeed, in which appeal is made to the "*prima facie*" truth of generalizations. I may suggest, incidentally, that it has relevance to a question raised earlier in this paper, namely, that of the relation between actions and bodily movements. My suggestion is that any bodily movement of a sort which is normally under voluntary control, i.e., excluding such movements as circulation of the blood or the reflex movement of limbs, is to be accounted an action in the absence of sufficient reasons to the contrary. Because it employs the notion of *sufficient reasons*, this suggestion is vague, but, I submit, it is not nearly so vague as the action analyst's suggestion that a bodily movement "under appropriate circumstances" is an action. When we speak of sufficient reasons to the contrary, we know at least the sort of circumstances to look for (the unseen push, the hidden wire, the unnoticed injection).

The reasons which defeat a given *prima facie* claim cannot be specified in a general and non-trivial way. As sufficient reasons defeating a given claim, a decision as to their presence, like the recognition of sufficient reason generally, involves an element of human judgment. And this is a defect only insofar as there is a reasonable expectation that there should be a purely mechanical test available for the truth of the sort of claim which may be in question.

My general claim, it may be recalled, is that in terms of our ordinary criteria, in the absence of defeating conditions, every action is one for which the agent is responsible. We may say that the fact that an agent has performed an action is a *reason* for holding him responsible for it. The question of whether or not an agent is to be held responsible for a given action will then be viewed as a question of balancing reasons *pro* and *contra*, with the initial presumption falling on the former side. Absolving conditions are, of course, of many sorts. If, e.g., *A* has stabbed *B*, *A* may be absolved of responsibility for his act by reason of the fact that he was acting under the orders of a military superior, or of the fact that *C* grasped *A*'s hand and pushed it toward

A's body. (In this case, the very claim that *A* stabbed *B* could be defeated.) Again, suppose that *D* has disclosed military information to the enemy—but only under extreme torture, or under the influence of a mysterious truth serum. Threats or physical abuse applied to one's self or to others, ignorance, illness, all these and other circumstances *may* serve to absolve an agent of responsibility for what he has done. But it must be borne in mind that each of the foregoing is but a general heading under which certain absolving conditions—which are always specific circumstances attending specific actions—fall. It is by no means the case that *any* circumstance falling under one of these headings is *ipso facto* an absolving condition—or even an extenuating circumstance.

Still, it is tempting to generalize about absolving conditions; to say, for example, that such conditions obtain with respect to a given action just in case the agent's performance of the action was unavoidable. On this account, "the agent couldn't have done otherwise" is taken as the general formula for an absolving condition. Such a formula is not so much wrong as unhelpfully general. Even granting that every instance of absolving conditions can be brought under the unavoidability formula, we find that application of this formula occurs in many different ways. Sometimes, e.g., "he couldn't have done otherwise" confers to nothing more than "it would have been unreasonable to have expected him to do otherwise." In most cases the applicability of the unavoidability formula requires as much *judgment* as does the determination that a given circumstance counts as an absolving condition. The unavoidability formula, therefore, would appear to be of little practical value in deciding whether or not an agent is responsible for a given action. It must be remembered, however, that what devices people may find useful in making judgments cannot be determined *a priori*.

However practically useful or useless the unavoidability formula may be in assisting persons to make judgments about absolving conditions, it is of central importance in the development of the Dilemma of Responsibility. For the rest of this paper I shall deal only with the view of determinism. I shall do so not because I think determinism is true—I have no idea to what extent events in the world may be subsumable under causal laws—but because I find the appeal to uncaused events completely unhelpful with respect to our philosophical puzzle. In any case, the philosophic—as opposed to the scientific—considerations in favor of in-

determinism are the alleged inadequacies of determinism. One such inadequacy is the supposed incompatibility of determinism and responsibility. This is our present concern, and the unavoidability rubric forms an important element of it.

For the present, let us take unavoidability as the general heading under which all absolving conditions fall. One consequence of this decision is that the demonstration of unavoidability is sufficient to absolve an agent of responsibility for a given action. In other words, if an agent could not but have performed a given action, then he is not to be held responsible for performing it. This is a rather familiar claim; it is a form of the doctrine that "ought implies can." This doctrine has for many philosophers the status of a truism, but I suspect that it is only the doctrine's vagueness which enables it to maintain this status.

Given that unavoidability is a defeating condition of responsibility, it is an easy step to the claim that determinism undermines the possibility of holding persons responsible for their actions, since determinism implies the unavoidability of all actions. I.e., given the presence of causally sufficient conditions of an action, an agent cannot but perform it. One of the determinist's standard replies to this charge is that it involves a confusion of compulsion and causation, and that only the former is incompatible with human responsibility. This reply does have the merit of suggesting that "unavoidability" does not apply to a homogeneous set of cases. But it is not completely satisfactory. First, the reply is frequently based on an empiricist analysis of causation ("All that one observes is billiard ball *A* striking billiard ball *B*, followed by the latter's moving . . ."). One cannot help wondering what an analogous analysis of *compulsion* would be like—perhaps, "All that one observes is *A*'s holding a pistol to *B*'s head, followed by *B*'s handing his wallet to *A* . . ."! The reply, more importantly, is insufficient as applied to cases of the causation of human behavior. Having left an airplane without a parachute, a person cannot but fall to the ground along a certain trajectory. (The question of avoidability may arise because on landing the person has struck and killed a child.) The unavoidability of this behavior is not due to "compulsion." (Nor would it be helpful for the determinist to point out that the operative causal laws are "descriptive" and not "prescriptive"!.) That the behavior in question—falling to the ground, or, even, killing the child—is not truly an *action* is not fully relevant to the present point. It

may be well, nonetheless, to give an example which might more clearly be described as involving the causation of a human action. Suppose, then, that we have developed a truth serum which, unlike wine or even sodium pentothal, is completely effective. Then, in circumstances in which an agent would not have told the truth but for having been given the drug in question, his being given the drug would properly be accounted the cause of his telling the truth (disclosing military secrets, e.g., or exposing illegal activities in high places). Although it is not necessary to do so, we may add the supposition that the drug acts, say, by neutralizing a center in the brain which normally serves to inhibit a natural tendency to tell the truth: this would serve to account for the fact that in the case described there is an *action* which, in a quite straightforward sense, is *caused*.

It might be thought, incidentally, that taking the drug is an absolving condition only if the agent has been *compelled* to take it. But this is not so. If the agent has taken the drug without compulsion, but either unwittingly or in ignorance of the drug's effect, this *might* constitute an absolving condition. But, again it might not. The agent's ignorance may be culpable; he may, for example, have been unreasonably inattentive during a lecture on the drug.

This point may serve to remind us of some of the difficulties involved in setting down a simple formula for absolving conditions, e.g., "compulsion or ignorance." Just as an agent may be blameworthy for being in a state of ignorance, so he may be blameworthy for being in a position in which he may be *forced* to perform an action. Thus, if a soldier, out of cowardice, or even out of stupidity, permits himself to be captured by the enemy, then no matter how courageously he may resist threats and even torture before disclosing military information, he is not entirely blameless for this action. In addition, the notion of compulsion is terribly vague. There is often a wide area of possible disagreement as to whether or not a person was *compelled* to perform a given action, turning on the question of whether he could *reasonably* have been expected, given the forces operating on him, to refrain from performing the action. And this is a question to be resolved, if at all, not by rule or formula, but by judgment.

. If unavoidability is to do the job of covering all cases of absolving conditions, then it must fit some cases of compulsion, of ignorance, and of causation of actions. But it cannot be assumed that un-

avoidability as an absolving condition applies to all cases falling into *any* of these three categories. Alternatively, if falling into one of these categories *ipso facto* renders an action *unavoidable*, then on this criterion of unavoidability, unavoidability is not necessarily an absolving condition. Herein, I submit, lies the basis for the resolution of our problem of responsibility and the causation of actions.

Continuing to assume the point of view of determinism, we may suppose, with one tradition within that general framework, that desires are causes of actions. This supposition comes to this: given certain circumstances, including an agent's perception of his situation, as well, normally, as his beliefs about the nature of a contemplated act and his expectations concerning its effects, the agent's desire to perform the action is a sufficient—and indeed also necessary—condition of his doing so. Like other generalizations about causes of events of a certain sort, this statement needs to be qualified—by reference, e.g., to the action's being within the power of the agent, to the agent's knowing what he is doing, etc. But these qualifications do not affect the general supposition of desires as causes of actions.

The action in question may be impulsive or carefully reasoned. In the former sort of case, the answer to the question "Why did you do so-and-so?" may be simply, "Because I wanted to." In the latter sort of case, the corresponding answer will be given in terms of some feature(s) of the act or of its expected consequences. Every action has features and consequences not anticipated by the agent. Described in terms of one of these features or consequences, the action is unintentional. That an action is unintentional, incidentally, does not necessarily constitute an absolving condition. Culpable ignorance may be involved, i.e., the agent *should have* been aware that the action would have the feature or consequence in question.

On our schema, desires are often efficacious in the production of actions; whether they are properly labeled "causes" becomes a merely verbal matter. If we deny the efficacy of desires, then we are hard put to account for their value in the explanation of actions. But worse, if, with the action analyst, we deny the causal efficacy of desires, then, as I have argued, the burden of the causal explanation of the bodily movements which constitute actions falls entirely on the causal antecedents of these bodily movements, e.g., an electrical discharge in the nervous system. Whatever the intention of the action analyst, his analysis

tends to relegate desires, beliefs, etc., to the role of epiphenomena. And epiphenomenalism is not only false to our experience, but it fails to account for both our ability to initiate actions and our responsibility for performing them.

But if desires are causes of actions, our old problem remains: in given circumstances, the occurrence of a desire to act appears to make the consequent action unavoidable, and this unavoidability seems to rule out the possibility of properly attributing responsibility to the agent. I have suggested that this difficulty can be challenged in either of two ways, by denying that causation entails unavoidability, or by denying that unavoidability is necessarily an absolving condition. Determinists have traditionally adopted the first alternative; the second strikes me as preferable.

There are cases in which we say of an agent that he could not help performing a given action. But the fact that the agent wanted to perform the action is not taken as grounds for such a statement. We don't ordinarily hold that someone's (non-obsessive) desire to perform an action renders its performance unavoidable. Confronted with this fact, the determinist has traditionally maintained that it shows that the causation of actions does not entail their unavoidability. In support of this point, the determinist has generally appealed to the claim that "could have done otherwise" has more than a single sense, or to a hypothetical analysis of this expression, which I shall discuss below. Now, as I noted above, the philosopher's appeal to different meanings of expressions raises serious difficulties. In any case, the claim that causation entails unavoidability itself makes a strong claim on common sense belief. It is simply not clear that the beliefs of common sense in this area are coherent, a fact strongly suggested, incidentally, by the persistence of philosophical perplexity here.

Instead of denying that causation entails unavoidability, we can question the claim that unavoidability is incompatible with responsibility. In other words, we can question the dogma that "ought implies can," or as Broad put it in a famous essay setting forth one version of the Dilemma of Responsibility,⁷ "obligability entails substitutability." Broad's claim is roughly equivalent, in the jargon of the present paper, to the claim that responsibility entails avoidability. Now the use of "entails" in this victim renders it too strong and,

hence, false. To be plausible, the relation between responsibility and avoidability must be set forth in a moral judgment, not in an analytic statement. It might seem plausible to hold that one ought not to hold an agent responsible for an action which he could not help performing. But, surely, to blame someone for an unavoidable action, while perhaps unfair, and even pointless, is not to be guilty of a logical inconsistency.

To be sure, one could so define "responsibility" as to render "responsibility entails avoidability" analytic. But to do so would rob the notion of responsibility of its normal use. In particular, the fundamental connection between responsibility and praiseworthiness or blameworthiness would be cut. For if this connection is maintained, then since praiseworthiness and blameworthiness are moral notions, and avoidability is not, to claim that responsibility entails avoidability is to commit a version of the naturalistic fallacy.

In questioning the dictum that ought implies can, it may be helpful to think of the familiar legal doctrine of strict liability, under which, in effect, the appeal to excusing conditions for certain classes of actions is legally disallowed. To be sure, application of this doctrine to some cases may be unfair, or even pointless, but its application is not always unwarranted. If the absentee owner of a milk processing plant is fined because the plant produces adulterated milk, this punishment may well lead him to take steps to insure that such an untoward occurrence is not repeated. The fact that under the circumstances as they existed, the owner of the plant might have been unable to affect the purity of the milk does not render his punishment pointless, or even necessarily unfair.

In a somewhat analogous way, we should be justified in holding individuals morally responsible for their actions even if the actions are caused and, therefore ("in a sense," if you like), unavoidable. The justification lies in the utility of the institution of moral responsibility as a device for the social direction of behavior. The point is not simply, as some determinists seem to suggest, that rewards and punishments serve as—respectively—positive and negative reinforcements of actions. Nor is it merely that, in addition, the application of sanctions, including praise and blame, provides examples whereby the behavior of other individuals is redirected or reinforced. Like the law, morality becomes institutionalized. Holding persons re-

⁷ "Determinism, Indeterminism, and Libertarianism" in *Ethics and the History of Philosophy* (London, 1952), pp. 195-217. See esp. pp. 196ff.

sponsible for their actions is an important part of the institution of morality (as it is of law). Knowing that agents, including ourselves, are held responsible for their actions is a significant factor influencing our perception of a given social situation, and of our desires to act in particular ways in the situation. Much more, of course, needs to be said about the workings of the practice of holding persons responsible for their actions. But if the existence of the practice serves to increase the incidence of socially useful behavior, then, in the absence of overriding considerations to the contrary, this fact alone justifies the practice. And holding persons morally responsible for their actions, as a *practice*, should demand nothing more than a pragmatic justification. My general point is that the fact that we justifiably hold persons responsible for their actions does not necessarily imply that, given the total situation in which they acted, they could have acted otherwise: the very fact that we hold persons responsible alters the situations in which persons act.

One aspect of this overly schematized account is that the causal efficacy of desires under given circumstances need not (and, of course, does not) render these desires absolving conditions. The refusal to count desires as absolving conditions is justified by the fact that this refusal may, by leading to an altered perception of the agent's situation, lead to an alteration in his desires, and thereby to an alteration in his behavior. In the simplest sort of case, the expectation of punishment for a given action which would otherwise have been desired may lead to the desire for, and the performance of, an alternative action.

Still, to revert to an earlier point, when a person performs an action because he wants to, we don't ordinarily describe the situation by saying that he couldn't help performing the action, but that he is nonetheless not to be absolved of responsibility for it; rather, we say that he could very well have done otherwise. This fact requires explanation. First, it should be noted that, except from the pulpit or the political podium, we don't ordinarily make philosophical pronouncements (e.g., "causes render actions unavoidable," "ought implies can") at all, nor do we think them or say them to ourselves. In any case, we must be aware of the *point* of saying, "He could have done otherwise." And the point of this utterance is normally just to make the claim that the individual in question is not, under the given circumstances, to be absolved of responsi-

bility for his action, just as the point of saying that someone could *not* help performing a given action is normally to try to absolve him of responsibility for the action. No particular philosophical analysis of these claims, I submit, need be implied by their specific formulation.

A common procedure of determinists—and I am here concerned only with those determinists who maintain that their doctrine is compatible with human responsibility—has been to try to incorporate the fact of usage indicated in the preceding paragraph into the analysis of avoidability. They have done this by proposing an analysis in which this concept is treated as essentially hypothetical—in which "*X* could have done otherwise" is equated with something like "*X* would have done otherwise if he had wanted to (i.e., if his desires had been different)." Criticisms of the *adequacy* of such an analysis are familiar;⁸ I am here concerned with the not unrelated question of the *relevance* of such an analysis to our problem. It will surely be admitted that if some of the circumstances in which an agent acted had been suitably altered, the action would have been different. But this fact is simply irrelevant to the question of whether, given the circumstances as they in fact were, the agent could have acted otherwise than he did. And *this* is the question raised when we ask about the avoidability of an action. Critics of the hypothetical analysis of avoidability who argue that "*X* could have done otherwise" has a categorical use point, at least indirectly, to this question of the relevance of the hypothetical analysis. The fact that the circumstances thought of as altered are desires, rather than some other (causal) factors, is not directly relevant to the question of avoidability, although generally it is relevant to that of responsibility.

When all this has been said, determinism will still seem to many as a doctrine whose truth would threaten their status as free and responsible agents. That all events, including human actions, occur as the result of antecedent causal conditions seems to make our actions little different from the movements of a gear in a clock or of a leaf blown by the wind. To counteract this picture, we must remember that the sort of determinism with which we are concerned posits as causes human beliefs and human desires. The fact that our desires are efficacious hardly implies a lack of freedom. To act in accordance with one's desires is not to be pushed or pulled—indeed, *not* to act in accordance with one's desires is at least the beginning of a

⁸ See, e.g., J. L. Austin's "Ifs and Cans," reprinted in *Philosophical Papers* (Oxford, 1961), pp. 153–180.

characterization of compulsion. To be sure, some desires become obsessive, but they are far from a representation of the standard case.

Determinism is appalling to some philosophers because acting in accordance with one's desires has a bad name among them. Sometimes this ill repute has its source in a view of man's nature as inherently sinful; in any case it construes the class of human desires far too narrowly. Surely, there are base desires. But there are also the quite human desires to know and to speak the truth, to help others, to create beautiful things, and even to perform actions because they are right. It is the function of moral training to inculcate desires such as these. Society's sanctions serve to remedy the failures and

deficiencies of moral training. Recognizing that men cannot help the nature of their moral training, we must nonetheless continue to hold them responsible for their actions, albeit with compassion.

There are many features of this world which are deplorable. But the fact—if it is a fact—that all events in the world fall under causal laws should not be accounted one of them. The world—just as it is—*may* be deterministic. It is only certain pictures of the world suggested by determinism which appall and, perhaps, terrify. But the view that our actions are the effects of our desires should not so affect us—or should we prefer a world in which our desires to act are without effect?

University of Washington

Received July 1, 1968

III. MINDS AND IDEAS IN BERKELEY

GEORGE PITCHER

IT is an important part of Berkeley's official doctrine that the mind is an immaterial, spiritual substance. I want to examine critically his views concerning the metaphysical relationship between this substance and its ideas. Berkeley devotes a lot of attention to the connection between ideas and such things as tables and chairs, but he is remarkably slapdash in his treatment of the connection between ideas and the mind that has them. I do not mean that he does not have strong opinions on the subject, for he does. But the opinions are only briefly and inadequately defended. And what is worse, they seem not to be consistent with one another.

Berkeley asserts each of the following propositions:

- (A) The mind perceives ideas.¹
- (B) The mind is wholly distinct from its ideas.
- (C) The alleged distinction between (i) the perceiving of an idea and (ii) the idea perceived, is a bogus one; there is no such distinction.

In this paper I shall do the following. First, I shall give textual justification for my claim that Berkeley did in fact hold each of the theses (A), (B), and (C). I shall then show that (A), (B), and (C) together constitute an inconsistent triad of propositions. Then, since (B) and (C) are perhaps the most suspicious-looking of the three, I shall give Berkeley's reasons for asserting them. Next, I shall assess the merits of these reasons, and finally, indicate the most plausible moves Berkeley might make to extricate himself from inconsistency.

No one will dispute that Berkeley subscribed to (A). Here are some representative passages:

Besides all that endless variety of ideas or objects of knowledge, there is likewise something which knows or perceives them and exercises divers operations, as willing, imagining, remembering, about them. This perceiving, active being is what I call "mind," "spirit," "soul," or "myself." (*Of the Principles of Human Knowledge*—hereinafter *PHK*—Pt. I, sect. 2.)

A spirit is one simple, undivided, active being—as it perceives ideas it is called "the understanding" . . . (*PHK*, Pt. I, sect. 27.)

. . . I myself am not my ideas, but somewhat else, a thinking active principle that perceives, knows, wills, and operates about ideas. (*Three Dialogues between Hylas and Philonous*—hereinafter *3D*—third dialogue: A. A. Luce and T. E. Jessop (eds.), *The Works of George Berkeley*—hereinafter *Works*—II, 233.)

That Berkeley held (B) is almost as easy to verify. In addition to the passage just quoted in support of (A), one may also cite the following:

This perceiving, active being is what I call "mind," "spirit," "soul," or "myself." By which words I do not denote any one of my ideas, but *a thing entirely distinct from them*, wherein they exist or, which is the same thing, whereby they are perceived . . . (*PHK*, Pt. I, sect. 2. Italics mine.)

"Thing," or "being" is the most general name of all; it comprehends under it two kinds entirely distinct and heterogeneous, and which have nothing common but the name, to wit, *spirits* and *ideas*. (*PHK*, Pt. I, sect. 89.)

Spirits and *ideas* are things so wholly different that when we say "they exist," "they are known," or the like, these words must not be thought to signify anything common to both natures. There is nothing alike or common in them . . . (*PHK*, Pt. I, sect. 142.)

Ideas are things inactive, and perceived: and spirits a sort of beings altogether different from them. (*3D*, third dialogue: *Works*, II, 231.)

Proposition (C) is one that Berkeley asserts most emphatically in the *Philosophical Commentaries* (*PC*), where he writes:

The distinguishing betwixt an idea and perception of the idea has been one great cause of imagining material substances. (*PC*, entry 609.)

and

. . . wherein I pray you does the perception of white differ from white. (*PC*, entry 585.)

¹ It would be closer to ordinary usage to speak of the mind's *having* ideas, or being *aware* of ideas, or being *conscious* of ideas; but I shall follow Berkeley and speak of the mind's *perceiving* ideas.

It is not announced in such loud tones in the works meant for publication, but it does show up there too.

Light and colours, heat and cold, extension and figures—in a word, the things we see and feel—what are they but so many sensations, notions, ideas, or impressions on the sense? And is it possible to separate, even in thought, any of these from perception? For my part, I might as easily divide a thing from itself. (PHK, Pt. I, sect. 5.)²

The distinction is denied again in the *Three Dialogues*. In the first dialogue, Philonous (Berkeley) picks up a remark made earlier by Hylas:

To return then to your distinction between *sensation* and *object*; if I take you right, you distinguish in every perception two things, the one an action of the mind, the other not. (Works, II, 195.)

Philonous then proceeds to attack the distinction, on grounds that we shall be discussing in a moment. He thinks he shows the distinction to be illegitimate, and near the end of his argument he says:

Since therefore you are in the very perception of light and colours altogether passive, what is become of that action you were speaking of, as an ingredient in every sensation? (Works, II, 197.)

There can be no doubt, then, that Berkeley subscribes to all three propositions (A), (B), and (C). But it is not at all clear how (B), especially, is to be understood. This is due not only to the obscurity of the concepts of mind and of idea, but also to the obscurity of the concept of distinctness (or wholly-distinctness). The mind, according to Berkeley's official doctrine, is an unextended, immaterial, or spiritual substance. It is reasonably certain, I think, that whatever conception we have of such an entity must be based on an analogy with ordinary *physical* substances—with such things as the bodies of human beings, for example.³ The passages I have quoted above in support of the claim that Berkeley avowed (A) and (B)—and they are typical of his statements about minds and their ideas—strongly suggest that he was implicitly thinking of the relation between a mind and any one of its ideas as analogous to the relation between

Bill's body and, say, a tree, when Bill kicks the tree. In this example, we may detect two kinds of distinctness, which I shall call strong distinctness and weak distinctness. Bill is (wholly) distinct from the tree in an obvious sense. I shall say that Bill is strongly distinct from the tree. If Bill kicks the tree, there may be some point in saying that Bill is distinct from his act of kicking the tree. Bill is not the same thing as his acts or his states, and so, in a *recherché* sense, he is distinct from them. But they, unlike the tree, qualify his being, to employ the terminology of another day. He is to some extent necessarily, even if only temporarily, different because of them. So he is not distinct from them in the strong way he is distinct from the tree: I shall say that Bill is weakly distinct from his acts and states.

We shall have to examine (B) much more critically later, but for the moment I shall assume that it must be understood as saying that the mind is distinct from its ideas in a way analogous to that in which Bill is distinct from the tree—in other words, that the mind is strongly distinct from its ideas. On this interpretation, an idea is not an element in the ontological make-up of the mind. To be sure, Berkeley says, as we have seen, that ideas exist in the mind, but he adds that he means by this only that ideas are perceived by the mind (PHK, Pt I, sect. 2)—and that is consistent with maintaining that ideas are no part of the metaphysical make-up of minds.

If it is indeed in this way that (B) is to be understood, then (A), (B), and (C) form an inconsistent trio of doctrines. One of the things the mind does is perceive ideas (proposition (A)); so the perceiving of an idea is either an act or a state of mind. Therefore the mind is not strongly, but only weakly, distinct from the perceiving of the idea. However, since the perceiving of an idea cannot be distinguished from the idea perceived (in virtue of (C)), the idea, like the perceiving, is only weakly distinct from the mind. And this contradicts (B), which says that an idea is strongly distinct from the mind.⁴

One or more of the three propositions clearly has to go. It would be possible to allow Berkeley

² This passage may be construed, however, as asserting that the existence of an idea consists in its being perceived, and hence as not necessarily asserting (C).

³ Berkeley has a related point in mind when he writes: "Speech [is] metaphorical more than we imagine, insensible things and their modes, circumstances, etc., being expressed for the most part by words borrowed from things sensible. The reason's plain. Hence manifold mistakes." (PC, entry 176.)

⁴ S. A. Grave—and, to the best of my knowledge, only he—has noticed this inconsistency in Berkeley. See his article, "The Mind and Its Ideas: Some Problems in the Interpretation of Berkeley," *Australasian Journal of Philosophy*, vol. 42 (1964), pp. 199–210.

to retain (B) and (C), thus forcing him to abandon (A). On this view, the perceiving of an idea and the idea perceived would be identified (in virtue of (C)), and the unified idea-perception would be kept wholly, or strongly, distinct from the mind, from mental substance (in virtue of (B)). This would mean, of course, that the mind is no longer the perceiver of ideas—i.e., the mind is no longer “doing the perceiving”—and that proposition (A) is therefore rejected. Berkeley could hardly be expected to approve of this solution, since (A) seems to be central to his thinking. And it is, anyway, a profoundly unsatisfactory position to defend: for it makes the relationship between the mind and its idea unintelligible. It is not enough to be told that the mind and its ideas are totally distinct: we want to know how they are related. For example, what is it that makes a certain idea an idea belonging to just this mind, rather than to some other? Berkeley cannot answer that this mind is the one that directly *causes* the idea; for according to him, all our ideas of sense are caused directly by God’s mind, not by our own. And if we are forbidden to answer “This mind is the one that *perceives* the idea,” we seem to have no answer whatever. And so we do not understand what the relationship between a mind and its idea could possibly be on this view. I conclude, therefore, that proposition (A) cannot be rejected, and that either (B) or (C), or both, must be sacrificed.

We must cast suspicious looks at these two propositions, then. I shall begin with (C). In the passage in which Berkeley deals most explicitly with (C) (namely, in the first dialogue of *3D*, *Works*, II, 194–197), we can discern two arguments against the alleged distinction referred to in that proposition. The first is that if you distinguish two elements in every perception of an idea—the perceiving of the idea and the idea perceived—the first element would have to be a (mental) act, since the phrase “perceiving an idea” is simply short for the expression “*the act of perceiving an idea*.” But there is no such act, since we are entirely passive in perception: the only actions involved in seeing, for example, are such acts as opening one’s eyes, turning oneself and one’s eyes in the right direction, and so on. And none of these constitutes any part of perception itself—for *that* takes place only after these acts are performed, and when it does take place, we are totally passive, apprehending whatever patterns of light and color

happen to impress themselves on our consciousness from outside.⁵ Since there is no such thing as an act of perceiving an idea, the alleged distinction between that (nonexistent) act and the idea perceived must be rejected.

This first argument for abandoning the distinction seems to be an incredibly weak one. Berkeley makes it quite clear that if something a person does is to be counted an *act*, it must be directly caused by one of that person’s volitions. But this is far too strong a requirement to impose for the purpose at hand. The issue to be resolved is whether one can or cannot distinguish the perceiving of an idea from the idea perceived. Berkeley argues, in effect, that the distinction is allowable only if the former (i.e., the perceiving) is an act, that is, only if it is directly caused by a volition. Since perceiving is not so caused, the distinction is ruled out. But this requirement appears to rule out far too much. Imposing Berkeley’s stricture, one could argue that if a wind blows you down and you fall on an ant, there can be no distinction between your falling on the ant and the ant fallen upon, since your falling on the ant, unprecedented as it was by a volition, was not an act. And this seems absurd.

Berkeley is anxious to deny the distinction between the perceiving of an idea and the idea perceived because he thinks that if you detach them from one another, then although the perceiving will of course be something occurring in the mind, the object of the perceiving—the idea perceived—will have to be located outside the mind in some “unperceiving substance,” since all ideas are “visibly inactive” and cannot be granted the status of mental entities. This brings us to Berkeley’s second argument. He says that if we are to distinguish the perceiving of an idea from the idea perceived, we must do so for all kinds of ideas. But in the case of pain, it is absurd to suggest that the alleged object of the alleged act of perceiving might exist outside the mind in an unthinking substance. Berkeley is on stronger ground here, but still his argument is far from decisive; for it rests on the dubious assumption that if one grants the disputed distinction, then perceived ideas must be accorded extra-mental existence. Indeed, Berkeley’s own argument could be turned around by a defender of the distinction: such a person could urge that since the distinction is obviously legitimate and since the object of the perceiving (the idea per-

⁵ Berkeley discusses only sense perception in this passage, and hence only one limited class of ideas; but it would not be difficult to construct parallel arguments for other kinds of ideas—e.g., images.

ceived) is at least sometimes not an extra-mental entity (as, for example, in the case of pains or dreams), it is just a mistake to assume that the distinction entails the extra-mental existence of the object that is perceived. Here we may begin to suspect that there might be something odd about proposition (C) and the distinction with which it deals—or at least that our understanding of the distinction may not be as perfect as we had supposed. We shall have to return to it later.

But let us first cast a jaundiced eye on proposition (B). Berkeley thought he had proved conclusively that ordinary physical objects—things like tables and chairs—are nothing but families of ideas. This conclusion naturally made him disposed to think of ideas as being strongly distinct from mind. Physical objects, after all, are normally conceived of as standing over against the mind, as being what the mind is not. It would be ludicrous to suggest that tables and chairs—or elements of them—might actually constitute parts of the mind. So ordinary common sense would urge. But these considerations ought not necessarily to carry any weight with Berkeley the philosopher: no doubt they exert a strong psychological force impelling one in the direction of (B)—but that is not at all the same thing as a rational argument.

The ground Berkeley most often gives in support of (B) is that ideas are inactive and so must be strongly distinct from the mind, which is active. But this is a feeble excuse for accepting (B): as we saw, Berkeley takes pains to point out that perceiving an idea is not an act. Presumably, then, the perceiving of an idea is inactive, too, a mere passive process. And yet it is one of the things that goes on in the mind (see proposition (A)) and so is not strongly, but only weakly, distinct from the mind. So the mere fact that an idea is inactive is not enough to make it strongly distinct from the mind, as (B) asserts.

I think it possible that Berkeley's own terminology may have beguiled him into a favorable attitude toward (B)—I mean his speaking of the mind as *perceiving* ideas. The model or picture embodied in this way of talking is that of ordinary sense perception: just as when a person sees a tree, the tree is something strongly distinct from the person (at least, unreflective common sense views the matter in this way), so when a mind perceives an idea, the idea must be something strongly distinct from the mind. But of course we have no reason whatever for thinking that this perceptual

model for the mind's awareness of ideas is a suitable one, despite the fact that a great many good philosophers have employed it. On the contrary, it would not be difficult to think of reasons for viewing that model with great suspicion.

I want to hazard the guess that none of the foregoing considerations weighed as heavily with Berkeley in embracing (B) as this one: he thought that the only alternative to (B) was the Cartesian view that an idea is a mode of consciousness or mind, and he regarded that view as absurd. Descartes held that an idea is a mode of consciousness, a way for the mind to be conscious (and hence, for Descartes, to exist). Just as being spherical and blue are modes of extension, i.e., ways of being extended, so having an image of a blue sphere is a mode of consciousness, i.e., a way of being conscious. This conception denies that ideas are strongly distinct from the mind, and thus stands in stark opposition to proposition (B). As Descartes puts it:

I allow only so much difference between the soul and its ideas as there is between a piece of wax and the various shapes it can assume.⁶

Berkeley does not attack this Cartesian view head-on, but it is quite clear from some of the things he says that he believes it to be absurd: for he thinks the view entails that when a person has a sense-impression or an image of, say a blue sphere, that person's mind is then blue and spherical. Here are two passages in which this criticism is implicit:

That there is no substance wherein ideas can exist besides spirit, is to me evident. And that the objects immediately perceived are ideas, is on all hands agreed. And that sensible qualities are objects immediately perceived, no one can deny. It is therefore evident there can be no *substratum* of those qualities but spirit, in which they exist, not by way of mode or property, but as a thing perceived in that which perceives it. (3D, third dialogue: *Works*, II, 237.)

It may perhaps be objected that if extension and figure exist only in the mind, it follows that the mind is extended and figured, since extension is a mode or attribute which (to speak with the Schools) is predicated of the subject in which it exists. I answer, those qualities are in the mind only as they are perceived by it—that is, not by way of *mode* or *attribute*, but only by way of *idea*; and it no more follows that the soul or mind is extended, because extension exists in it alone, than it does that it is red or blue, because those colors are on all hands acknowledged to exist in it, and nowhere else. (PHK, Pt. I, sect. 49.)

⁶ Descartes, Letter No. 347, in *Descartes: Philosophical Writings*, tr. by E. Anscombe and P. T. Geach (Nelson, 1954), p. 288.

In these passages, Berkeley evidently assumes that if an idea is a mode of consciousness, the mind must actually have the properties that are contained (in whatever sense they are contained) in the idea. In the first of the two passages just quoted, Berkeley argues that the objects we immediately perceive are ideas, and (at least in sense perception) the objects we immediately perceive are sensible qualities. From this we are obviously meant to infer that sensible qualities *are* ideas. So when ideas are alleged to be modes of consciousness, sensible qualities are also necessarily accorded the same status. But if sensible qualities, such as blueness or sphericity, are modes of consciousness, then the mind must have those qualities—must be, for example, blue and spherical. Since this is an absurd consequence, the view that ideas are modes of consciousness must be rejected.⁷

Berkeley's attack on the Cartesian view of ideas looks terribly weak—and I think it really is. The Cartesian seems to have the following easy rejoinder: "When I say the idea of a blue sphere is a mode of consciousness, I am not committed to holding that blueness and sphericity are modes of consciousness, but only that the *awareness* of blueness and sphericity is a mode of consciousness. And this does not in the least entail that the mind itself is blue and spherical." This rejoinder seems eminently sensible, but I suspect that Berkeley may have thought he could not allow it. I think he may have thought that the rejoinder presupposes a distinction between the awareness of an idea (or, in his terminology, the perceiving of an idea) and the idea of which one is aware (the idea perceived). This distinction, as we know, is rejected in Berkeley's proposition (C). Since there is no such distinction, he may have reasoned, if the awareness of blue sphericity is a mode of consciousness, so also is the blue sphericity—for the two are identical. So the absurdity cannot be avoided: the rejoinder merely seemed to avoid it by appealing to a nonexistent distinction.

But the Cartesian view cannot be so easily upset. The rejoinder I put into the mouth of its defender in the last paragraph was unfortunately worded, for it did suggest that the view rests on the distinction that Berkeley wishes to reject. This is ironic, because the whole point of the Cartesian view, when seen aright, is that it denies precisely this distinction. What the Cartesian view urges, to express it in Berkeleyan terms and to confine it to just one kind of idea of sense, is that the proper

analysis for sentences of the form "Mind *M* perceives an idea of an *FG*," where *F* is a color term and *G* is a term designating a shape—for example

- (I) "Mind *M* perceives an idea of a blue sphere" —

is not

- (IA) $(\exists x)(\exists y)(x \text{ is the mind } M \ \& \ x \text{ perceives } y \ \& \ y \text{ is an idea of a blue sphere})$

but rather

- (IB) $(\exists x)(x \text{ is the mind } M \ \& \ x \text{ perceives in a blue spherical manner}), \text{ or}$
 $(\exists x)(x \text{ is the mind } M \ \& \ x \text{ perceives blue-spherically}).$

The second conjuncts in (IB) are to be regarded as asserting that the mind *M* is in a not-further-to-be-analyzed state of consciousness. In particular, the analysis (IB) denies that the idea of a blue sphere is an object of awareness, separable from the awareness of it, as analysis (IA) suggests. Of course, a mind that is aware blue-spherically has those properties before itself as "objects" of awareness, but these "objects" are subjective sense-contents, essential aspects of the state of consciousness itself. According to (IB), the awareness of an idea is no more distinguishable from the idea than the turning of a somersault is distinguishable from the somersault turned, or than the swimming of the breast-stroke is distinguishable from the breast-stroke swum. Those who are tempted to analyze (I) along the lines of (IA), the Cartesian would maintain, are unconsciously assimilating (I) to such sentences as "Bill kicks the tree" or "Jane upsets the vase," where there is an action (or at any rate a bodily movement) and a seemingly distinct object. But there is absolutely no need to do this; and it is in fact far better to assimilate (I) to such sentences as "Jones turns a somersault" or "Tarzan is swimming the breast-stroke." It is better because one thus avoids the positing of a class of objects—viz., ideas—that are metaphysically embarrassing. If only on grounds of parsimony, (IB) is preferable to (IA); both are committed to the existence of minds and both presuppose that minds can be in certain states (or perhaps (IA) presupposes rather that minds can perform actions), but (IB) requires nothing further, while (IA) requires the existence of another kind of object.

If someone should complain that while he understands the analysis (IA) well enough, he does not

⁷ Locke rejects the Cartesian view on similar grounds: see his *Examination of Malebranche*, sect. 39.

understand either version of (IB) at all, I would reply in the following way. Our conception of a spiritual substance, as I remarked earlier, must evidently be based on an analogy with material substance. A philosopher who wishes to defend the analysis (IA) is in effect urging that we use as our model, in construing sentences of the form (I), such sentences as "Bill kicks the tree"; and a philosopher who defends analysis (IB) is urging that we use instead such sentences as "Jones turns a somersault." As far as I can see, on the score of intelligibility there is no *prima facie* warrant for preferring one model, or analogy, over the other. It may turn out, when the analogies are developed, that our intellectual hold on one is firmer than it is on the other—but there is no legitimate reason to pre-judge the matter at the very outset.

Someone might object that on the Cartesian view, the perceiving of an idea is a *state* of consciousness, not an *act* of consciousness; and yet I have maintained that on that view, sentences of the form (I) should be assimilated to such sentences as "Jones turns a somersault" and "Tarzan is swimming the breast-stroke," both of which clearly attribute *actions* (or perhaps in the case of swimming, it is rather an *activity*) to an agent. This is not a formidable objection, however. The essence of the Cartesian view⁸ is not that the mind's perceiving an idea is necessarily a *state*, as opposed to an *action*, of the mind; it is that the words "an idea" in the expression "the mind perceives an idea" do not designate an object distinct from the perceiving. It does not matter here whether the perceiving is an act, a state, a process, or something else; rather, the words "an idea" designate a way the mind perceives—whether perceiving is an act, a state, a process, or whatever it is. If we think an idea *must* be an object distinct from the perceiving of it, then we are being misled by a model suggested to us by the form of words we use. To rid ourselves of the model, we need only stop thinking of perceiving ideas as being at all analogous to kicking trees and upsetting vases, and start thinking of it instead as being analogous to dancing jigs and swimming breast-strokes. The words "a jig" in the expression "dancing a jig" do not designate an object distinct

from the dancing of it, whether the dancing is an act, a state, a process, or whatever it is;⁹ rather, the words "a jig" designate a way of dancing, whether the dancing is an act, a state, a process, or whatever it is. (If the objector insists on holding out for an analogue from the class of what are undoubtedly *states*, as opposed to acts or activities, then the Cartesian could reply that although suitable examples may be hard to come by in the real world, it is easy to find them in the realm of fantasy. Thus, imagine that we spoke not of sitting in a lotus position, but rather—as we easily might—of sitting a lotus.)

Notice how plausible the Cartesian view is in the case of sensations: for example, it certainly makes sense to deny that a pain is one thing and the feeling or awareness of it, another. There is just the feeling of pain. So accepting the Cartesian view for the whole range of ideas is tantamount to assimilating our awareness of all ideas to our awareness of sensations.

(IA) represents what we may call, adapting the terminology of modern sense-datum theories, an *act-object analysis*¹⁰ of the perceiving of ideas of sense, while (IB) represents an *adverbial analysis* of such perceiving. Looking back to our original triad of Berkeleian propositions, we see now that (IA) finds expression in proposition (B), and (IB) finds expression in (C). It is just this fact that accounts for the inconsistency in that triad of propositions: Berkeley shifts from an act-object analysis of perceiving ideas in (B) to an adverbial one in (C). This shift sets up an unresolved conflict that runs all through Berkeley's philosophy. I do not mean that a philosopher could not consistently accept an adverbial analysis for the perceiving of some ideas (e.g., pains) and an act-object analysis for the perceiving of other ideas (e.g., ideas received in sense perception). But there is no indication that Berkeley thinks our perceivings of ideas may be of two different sorts, and every indication that he thinks all ideas enjoy the same metaphysical status. His shifting back and forth between the adverbial and the act-object points of view is not governed by any explicit or implicit rules that I can discover; so I can only judge that he means to

⁸ I mean the view I am in the process of defending: I do not pretend that it corresponds in all respects with Descartes' actual view.

⁹ It is easy enough, by the way, to think of cases of jig-dancing, somersault-turning, and breast-stroke-swimming that are not *actions*, at any rate. For example, suppose that Jones, while standing, is suddenly rendered unconscious and that in falling, he turns a perfect somersault. Similarly, imagine a floating, but comatose Tarzan studded with electrodes that are able to inflict muscle-activating impulses; when Jane operates the impulse generator, Tarzan swims a jerky, but passable, breast-stroke.

¹⁰ Perhaps "passive state-object analysis" would be preferable to "act-object analysis" from Berkeley's point of view, but I shall nevertheless use the latter in what follows.

have just *one* view about our awareness of ideas. Since he does not, I say he falls into inconsistency.

I have been examining Berkeley's reasons for embracing propositions (B) and (C), and have found them wanting. If my criticisms of these reasons have been well taken, and if they are, as I believe, his *only* reasons, then Berkeley has provided no good ground for accepting either of those propositions. But there may well *be* such grounds—especially so if one wishes to defend the general metaphysical position that Berkeley does. So there may be good reasons why Berkeley *should* clasp either (B) or (C) to his bosom, despite the fact that he himself does not put them forward nor, presumably, realize what they are. And I think that in fact there are good reasons of this sort. In the concluding part of this paper, I shall try to show which of the two propositions Berkeley had better accept, and why—that is, which of the two leads to fewer and less embarrassing difficulties, and/or is consistent with more that is essential in his metaphysical system. To do this, I shall take a critical look at the distinction denied in (C); doing that will also, necessarily, be taking a critical look at proposition (B), for as we now realize, (B) affirms, at least, the very distinction that (C) denies.

I shall assume that Berkeley, insofar as he subscribes to proposition (C), is advocating an adverbial analysis of the perceiving of ideas. I shall also assume that insofar as he subscribes to proposition (B) he is advocating an act-object analysis. There may be more than the act-object analysis in (B), but it surely includes at least that. Rather than discuss (B) and (C) directly, then, I shall examine the act-object and adverbial analyses of the perceiving of ideas. I want to ask what someone who accepts the act-object analysis believes, or is committed to, that someone who accepts the adverbial analysis denies, or at least is not committed to.

Here is one answer to this question that looks promising: a defender of the act-object analysis maintains that although a given idea I_1 may in

fact be the object of an act (or state) of perceiving P_1 occurring at time t_1 , the same idea I_1 could exist also as the object of a different act (or state) of perceiving P_2 occurring at a different time t_2 . The words "the same idea I_1 " in this answer cannot be construed as meaning *generically* the same idea I_1 , because then, apart from the superficial matter of his expressing it in his own different terminology, a defender of the adverbial analysis would presumably agree with the proposition. He would agree, for example, that a mind can perceive blue-spherically at t_1 and blue-spherically at t_2 ; and doing this would be perceiving the generically same idea at two different times. In order to have any chance of marking a real difference between the act-object and the adverbial analysis, then, our answer must be interpreted so that "the same idea I_1 " means *numerically* the same idea. Now we *seem*, at least, to have something that an act-object analyst would say "Yes" to, and that an adverbial analyst would say "No" to. But this seeming is, as yet, deceptive. For there are two plausible choices still open to the act-object analyst. He can say either (a) ideas exist only when they are perceived, or (b) ideas exist when they are perceived and also when they are not perceived.

Let us explore the first possibility, (a). This proposition may be intended either as a necessary truth or as a contingent one. If the act-object analyst construes it in the first way, as a necessary truth, then, as far as I can tell, he is adopting a view that is indistinguishable from the adverbial analysis. For what is the difference between saying, as the adverbial analyst does, that sense-contents exist, of necessity, only when a mind is aware in a certain mode, and saying, as the act-object analyst does if he accepts (a) as a necessary truth, that ideas exist, of necessity, only when a mind perceives them? This is a mere seeming-difference; the act-object analyst's "object" is nothing but the adverbial analyst's sense-contents with a different label.¹¹

The only hope for an act-object analyst to keep

¹¹ This is not quite right: to put the point precisely, I should say that there is no difference if the act-object analyst claims that (a) is necessarily true *and, in addition, holds that no idea-residues exist when the corresponding ideas are not being perceived*. For the act-object analyst can maintain that the relationship between a mind and any one of its ideas is analogous to that between a husband and a wife: the two are strongly distinct, and yet there is a logical connection between them, since there can be no wife without a husband. But when a woman's husband dies, she does not need to cease existing: she may continue to exist, not of course as a wife, but as a woman, or, as we may put it, as a *wife-residue*. Similarly, the act-object analyst can hold that it is a necessary truth that ideas *as such* exist only when perceived, but maintain that they exist as *idea-residues* when they are not being perceived. (Cf. the early Russell on sense-data and sensibilia.) If the act-object analyst takes this line, there certainly is an important difference between his view and that of the adverbial analyst. Therefore, to be precise, I must add the second requirement specified (and italicized) in the first sentence of this note. In the text, I ignore this option that is open to the act-object analyst; because for our present purposes, (b), above, represents essentially the same option.

his analysis really distinct from the adverbial analysis is for him to deny that (a) is a necessary truth. But then he owes us an explanation of why an idea, although it could well exist when not being perceived, never in fact does so, of why ideas pop into, and out of, existence as minds start, and stop, perceiving them. If he were to give such an explanation, the act-object analyst would then have endowed ideas with an extremely thin, barely perceptible, margin of object-hood over the subjective sense-contents of his adverbial analyst opponent; but it is unlikely that the offered explanation would be anything but wildly implausible and unconvincing—which explains, I suppose, why no act-object analyst (as far as I know) has ever tried to formulate one. So if an act-object analyst is going to deny that (a) is a necessary truth—as he must if his analysis is to differ from the adverbial analysis—then he would be well advised not to hold on to (a) as a contingent truth but rather to reject it altogether and to accept (b) instead. For (b) accords ideas exactly the same metaphysical status that (a) does, when (a) is understood as a contingent proposition, but is simpler and more plausible.

If I am right, then, the real difference between the adverbial analysis and the act-object analysis is this: on the adverbial analysis, it is a necessary truth that no ideas can exist unperceived, while on the act-object analysis, ideas can and do exist unperceived. Before drawing what are doubtless the obvious conclusions from all this, I should like to point out that Berkeley agrees with me that any philosopher who denies proposition (C) (which embodies the adverbial analysis of the perception of ideas) and thus accepts what I have been calling the act-object analysis, is committed to the existence of unperceived ideas.

‘Twas the opinion that ideas could exist unperceived or before perception that made men think perception was somewhat different from the idea perceived . . . (PC, entry 656.)

I cannot undertake here to discuss the relative merits of the adverbial and act-object analyses of the perception of ideas, nor is that any part of my purpose in this paper. I want only to extricate Berkeley, as painlessly as possible, from what I take to be a serious inconsistency. It should now be fairly clear how this can be done. Proposition (B) embodies the act-object analysis which entails, if

I am right, that ideas can exist when unperceived; proposition (C) embodies the adverbial analysis which entails that it is necessarily false that ideas exist when unperceived. Obviously, then, the simplest and best way to preserve the largest bulk of Berkeley’s central philosophical doctrines is to have him abandon proposition (B). And he would lose nothing of importance by doing this; for the characteristics he was anxious to attribute to ideas—passivity above all else—he could now simply attribute to the “sense-contents” of the adverbial analysis. Indeed, it seems easier to understand why such sense-contents should be inactive than why ideas, conceived as objects, should be so.

There is at least one famous Berkeley scholar, namely, the Rev. Canon A. A. Luce, who would violently object to my suggestion. He does not think that Berkeley is guilty of any inconsistency: ignoring, or at any rate attaching no importance to, the passages I have cited above in support of proposition (C), Luce maintains that Berkeley simply holds propositions (A) and (B). On this view, ideas—or rather, just ideas of *sense*—are objects that are sharply to be distinguished (i.e., are strongly distinct) from anything mental:

Such ideas are not mental . . . The Berkeleian idea of sense is *from* and *in* the mind of God, and *for* and (sometimes) *in* the mind of man, but it is not mental. It is not a constituent of mind, divine or human. It is the non-mental *other* of mind.¹²

This interpretation has Berkeley accepting the act-object analysis of ideas of sense, and spurning the adverbial analysis that I have urged on him.

If Berkeley really does think of ideas as objects, in the way Luce suggests, then if what I have said about the difference between the act-object and the adverbial analyses is correct, Berkeley must hold that it is logically possible for ideas¹³ to exist unperceived. He then has the further option of saying that (i) ideas do in fact exist unperceived, or (ii) ideas in fact exist only when perceived. I think it is a fatal, or near-fatal, objection to Luce’s account that Berkeley most certainly does not think it logically possible that ideas might exist unperceived.

Luce might try to get around this objection by arguing that Berkeley’s reasons for thinking that it is not logically possible for ideas to exist unperceived are faulty, and that therefore he ought not to have asserted any such thing. So we ought to

¹² A. A. Luce, “The Berkeleian Idea of Sense,” *Proceedings of the Aristotelian Society*, Supplementary Volume XXVII (1953), p.3.

¹³ In what follows, by “idea” I shall mean idea of *sense*.

patch up Berkeley's philosophy by deleting the doctrine that it is logically necessary that ideas exist only when perceived: the way would then be open for accepting Luce's account.

I agree that Berkeley's grounds for holding the offending doctrine are bad ones. His main reason for accepting it is that he thinks he can show that "to exist," in the case of ideas, just *means* "to be perceived." This is not a good reason, for Berkeley's demonstration is invalid: but we cannot deprive Berkeley of it without robbing him of his most prized possession. He regards his thesis about what "existence" means as one of his greatest discoveries. Indeed, he states at least twice that he is entirely willing to rest his whole case for immaterialism on the thesis. (See *PHK*, Pt I, sect. 22, and *3D*, first dialogue: *Works*, II, 200. See also *PC*, entry 491.) So, to delete this proposition from Berkeley's philosophy, as the Luce tactic would require, would be to perform drastic surgery.

But suppose we overlook this huge difficulty in Luce's objection-avoiding move; and suppose we accept his interpretation of Berkeley entirely. This means, as I have pointed out, that Berkeley must hold that it is logically possible for ideas to exist unperceived. He must reject out of hand the further option (i)—viz., the view that ideas do in fact exist unperceived—and so must accept option (ii)—viz., the view that ideas in fact exist only when perceived. But now he is faced with what looks like an insuperable difficulty: for now he must give us some account, some explanation, of *why* these objects (i.e., ideas) happen to exist only when they are perceived by some mind or other—and it seems extremely unlikely that any plausible account of this kind could possibly be formulated. Certainly Berkeley makes no attempt to provide one. And incidentally, Luce must agree with me that some such account is really required on his reading of Berkeley, for he argues that Berkeley's ideas (of sense) are public objects, accessible to any number of different observers¹⁴—and surely anyone would have to admit that if such objects exist only when they are perceived, this is an extraordinary fact that cries out for some explanation.

I conclude that Luce's "act-object" interpretation of Berkeley saddles our philosopher with a hopeless position. I do not for a moment dispute that Luce has correctly spotted one central strand in Berkeley's thought—the strand, namely, ex-

pressed in proposition (B). I also acknowledge that he has seen some of its implications. But he has not realized that other implications of it clash with doctrines that Berkeley holds dear. To remedy this situation, I am suggesting that we allow Berkeley to abandon proposition (B), with its act-object analysis of the perceiving of ideas, and have him adopt instead the adverbial analysis implicit in proposition (C). This move has the great advantage of allowing him to retain the necessary non-existence of unperceived ideas as part of his system. Furthermore, everything else of importance that Berkeley wishes to maintain can be accommodated within the adverbial analysis.

This latter claim needs an extended defense, which I have no space to give it here. But I have already said a word or two in its favor, and I shall conclude by taking up two additional examples. The first is this: on the adverbial analysis, there is a sense in which two or more people may be said to perceive the same object. Since the sense-contents of the adverbial analysis are private to each person, it seems at first glance as though no two people could ever perceive the same objects (i.e., houses, chairs, tables). If so, this would constitute an objection to the analysis. Hylas makes precisely this point:

But the same idea which is in my mind, cannot be in yours, or in any other mind. Doth it not therefore follow from your principles, that no two can see the same thing? And is not this highly absurd? (*3D*, third dialogue: *Works*, II, 247.)

But the adverbial analyst is ready with an answer: As long as the sense-contents of two or more people agree to a suitable extent and in familiar ways—something that will be revealed, presumably, in their words and actions—then it is perfectly proper to say, as the vulgar do, that those people are seeing (hearing, touching, or whatever) the same things. (And this is just the line that Berkeley in fact takes in answering Hylas.¹⁵) The adverbial analysis allows the common ways of speaking to be preserved; it just gives those sentences of ordinary language learned interpretations. "In such things," as Berkeley pointed out in a famous passage, "we ought to 'think with the learned and speak with the vulgar'." (*PHK*, Pt. I, sect. 51.)

The second example is this: On the adverbial analysis, Berkeley can even insist, as he is wont to

¹⁴ See Luce, *op. cit.*, pp. 17-18.

¹⁵ See the criticism of Berkeley's position here in David Braybrooke, "Berkeley on the Numerical Identity of Ideas," *Philosophical Review*, vol. 64 (1955), pp. 631-636.

do, that the mind is (wholly) distinct from its ideas—not, of course, strongly distinct, but weakly so. The very words of proposition (B) may thus be preserved, as long as we understand by “distinct,” *weakly distinct*. There can be little doubt that when Berkeley talks in the (B)-vein, he implicitly thinks of

the mind as being strongly distinct from its ideas. On this point, I am in complete agreement with Luce. But if we understand weak distinctness where Berkeley intends strong distinctness, we can free him at last from the charge of inconsistency.¹⁶

Princeton University

Received September 9, 1968

¹⁶ I read this paper at Drew University, the University of Cincinnati, Johns Hopkins University, and the University of Minnesota. Comments and criticisms from philosophers at those institutions have helped me to remedy various defects in earlier versions of the paper. I am also grateful to my colleagues Richard Rorty and James Ward Smith for reading and commenting on the first draft of the paper.

IV. OMNIPOTENCE AND GOD'S ABILITY TO SIN

NELSON PIKE

IN the first chapter of the *Epistle of James* (verse 13) it is said that "God cannot be tempted by evil." This idea recurs in the confessional literature of the Christian tradition,¹ and is stated in its fullest form in the theological doctrine of God's *impeccability*.² God is not only free from sin, He is incapable of moral deviation. God not only does not sin, He *cannot* sin. This is generally held to be part of what is communicated in the claim that God is perfectly good. On the surface, at least, this doctrine appears to be in conflict with the traditional Christian doctrine of divine omnipotence. An omnipotent being is one that can do all things possible. But, surely, it is possible to sin. Men do this sort of thing all the time. It would thus appear that if God is perfectly good (and thus impeccable) He cannot sin; and if God is omnipotent (and thus can do all things possible), He can sin.

This argument appears to be sophistical. We are tempted to dismiss it with a single comment, viz., it involves an equivocation on the modal element in the statement "God can (cannot) sin." In the long run, I think (and shall try to show) that this single remark is correct. But that's in the long run; and in the interim there is a complicated and interesting terrain that has not yet been adequately explored. In this paper I shall discuss this matter in detail. After working through what I judge to be a number of conceptual tangles that have accumulated in this literature on this topic, I shall end by making a suggestion as to how the various senses of "God can (cannot) sin" ought to be sorted out.

I

I shall begin by identifying three assumptions that will work importantly in the discussion to follow.

First, I shall assume that within the discourse of the Christian religion, the term "God" is a descriptive expression having an identifiable meaning.

It is not, e.g., a proper name. As part of this first assumption, I shall suppose, further, that "God" is a very special type of descriptive expression—what I shall call a *title*. A title is a term used to mark a certain position or value-status as does, e.g., "Caesar" in the sentence "Hadrian is Caesar." To say that Hadrian is Caesar is to say that Hadrian occupies a certain governmental position; more specifically, it is to say that Hadrian is Emperor of Rome. To affirm of some individual that He is God is to affirm that that individual occupies some special position (e.g., that He is Ruler of the Universe) or that that individual has some special value-status (e.g., that He is a being a greater than which cannot be conceived).

Secondly, I shall assume that whatever the particular semantical import of the term "God" may be (i.e., whether it means, for instance, "Ruler of the Universe," "a being than which no greater can be conceived," etc.), the attribute-terms "perfectly good," "omnipotent," "omniscient," and the like, attach to it in such a way as to make the functions "If x is God, then x is perfectly good," "If x is God, then x is omnipotent," etc., necessary truths. It is a logically necessary condition of bearing the title "God," that an individual be perfectly good, omnipotent, omniscient, and so on for all of the standard attributes traditionally assigned to the Christian God. If we could assume that in order to be Emperor (as opposed to Emperess) of Rome one had to be male (rather than female), then if " x is Caesar" means " x is Emperor of Rome," then "If x is Caesar, then x is male" would have the same logical status as I am assuming for "If x is God, then x is perfectly good," "If x is God, then x is omnipotent," etc.

If there is an individual (e.g., Yahweh) who occupies the position or has the value-status marked by the term "God," then that individual is perfectly good, omnipotent, omniscient, etc. If He were not, then He could not (logically) occupy the position or have the value-status in question. However, with respect to the predicate "perfectly

¹ See, for example, the *Westminster Confession*, ch. V, sect. IV and the *Longer Catechism of the Eastern Church*, sects. 156–157.

² See the *Catholic Encyclopedia* (New York, Robert Appleton Co., 1967).

good," I shall assume that any individual possessing the attribute named by this phrase might not (logically) have possessed that attribute. This assumption entails that any individual who occupies the position or who has the value-status indicated by the term "God" might not (logically) have held that position or had that status. It should be noticed that this third assumption covers only a *logical* possibility. I am not assuming that there is any real (i.e., material) possibility that Yahweh (if He exists) is not perfectly good. I am assuming only that the hypothetical function "If x is *Yahweh*, then x is perfectly good" differs from the hypothetical function "If x is *God*, then x is perfectly good" in that the former, unlike the latter, does not formulate a necessary truth. With Job, one might at least *entertain* the idea that Yahweh is not perfectly good. This is at least a *consistent* conjecture even though to assert such a thing would be to deny a well-established part of the Faith.³

I now want to make two further preliminary comments—one about the predicate "omnipotent" and one about the concept of moral responsibility.

Pre-analytically, to say that a given individual is omnipotent is to say that that individual has unlimited power. This is usually expressed in religious discourse with the phrase "infinite power." St. Thomas explicated the intuitive content of this idea as follows: "God is called omnipotent because He can do all things that are possible absolutely."⁴ As traditionally understood, St. Thomas' formula must be given a relatively restricted interpretation. The permissive verb "do" in "do all things possible" is usually replaced with one of a range of more specific verbs such as "create," "bring about," "effect," "make-to-be," "produce," etc.⁵ God's omnipotence is thus to be thought of as creative-power only. It is not to be understood as the ability to *do* anything at all, e.g., it is not to be interpreted as including the ability to swim the English Channel or ride a bicycle. God is omnipotent in that He can create, bring about, effect, make-to-be, produce, etc., anything possible absolutely. For St. Thomas, something is "possible absolutely" when its des-

cription is logically consistent. Thus, on the finished analysis, God is omnipotent insofar as He can bring about any consistently describable object or state of affairs. In his article on "omnipotence" in the *Catholic Encyclopedia*,⁶ J. A. McHugh analyzes the notion in this way. It seems clear from the context of this piece that McHugh meant to be reformulating St. Thomas' view of the matter. I might add that I think this restricted interpretation of the pre-analytical notion of infinite power is an accurate portrayal of the way this concept works in the ordinary as well as in most of the technical (theological) discourse of the Christian religion.

Now, let us suppose that an innocent child suffers a slow and torturous death by starvation. Let it be true that this event was avoidable and that no greater good was served by its occurrence. Let it also be true that neither the child (or its parents) committed an offense for which it (or its parents) could be righteously punished. This is a consistently describable state of affairs (whether or not it ever occurred). I think it is clear that an individual that knowingly brought this state of affairs about would be morally reprehensible.

We can now formulate the problem under discussion in this paper more rigorously than above. God is omnipotent. When read hypothetically, this statement formulates a necessary truth. On the analysis of "omnipotent" with which we are working, it follows that God (if He exists) can bring about any consistently describable state of affairs. However, God is perfectly good. Again, when read hypothetically, this statement formulates a necessary truth. Further, an individual would not qualify as perfectly good if he were to act in a morally reprehensible way. Thus, the statement "God acts in a morally reprehensible way" is logically incoherent. This is to say that "God sins" is a logical contradiction.⁷ Hence, some consistently describable states of affairs are such that God (being perfectly good) could not bring them about.⁸ The problem, then, is this: If God is both omnipotent and perfectly good, there are at least some consistently describable states of affairs that He both

³ The truth of this assumption is argued at some length by C. B. Martin in the fourth chapter of his *Religious Belief* (Ithaca, Cornell Press, 1964).

⁴ *Summa Theologica*, Pt. I, Q. 25, a 3. This passage taken from *The Basic Writings of St. Thomas Aquinas*, ed. by Anton Pegis, p. 263.

⁵ These verbs are sometimes called "factive verbs."

⁶ New York, Robert Appleton Co., 1911.

⁷ There is probably some distinction to be made between acting in a morally reprehensible way and sinning. However, for purposes of this discussion, I shall treat these concepts as one.

⁸ I am here assuming that if God brings about a given circumstance, He does so *knowingly*. God could not bring about a given circumstance by mistake. I think this follows the idea that God is omniscient.

can and cannot bring about. There would thus appear to be a logical conflict in the claim that God is both omnipotent and perfectly good.

I think it is worth noting that the problem just exposed is not the same as the classical theological problem of evil. The problem of evil is generally formulated as follows: Evil exists. If God exists and is omnipotent, He could have prevented evil if He had wanted to. If God exists and is perfectly good, He would have wanted to. Since evil in fact exists, it follows that God does not exist. This argument is supposed to point up a conflict between the attribute of perfect goodness and the attribute of omnipotence. But the conflict is not of a rigorous sort. So far as this argument goes, it is logically possible for there to exist a being who is both perfectly good and omnipotent. The argument is supposed to show only that since it is contingently true that evil exists, it is contingently false that omnipotence and perfect goodness are possessed by a single individual. However, the problem we are now discussing has a sharper report than this. The argument generating this latter is supposed to show that there is a direct logical conflict between the attribute of perfect goodness and the attribute of omnipotence. No contingent premiss is employed (such as, e.g., that evil exists) and the conclusion drawn is that it is logically impossible (not just contingently false) that there exists an individual who is both omnipotent and perfectly good.

II

In reply to objection 2, article 3, question 25, Part I of the *Summa Theologica*, St. Thomas Aquinas writes as follows⁹

To sin is to fall short of a perfect action; hence to be able to sin is to be able to fall short in action, which is repugnant to omnipotence. Therefore, it is that God cannot sin, because of his omnipotence. Now, it is true that the philosopher says that *God can deliberately do what is evil*. But this must be understood either on condition, the antecedent of which is impossible—as, for instance, if we were to say that God can do evil things if He will. For there is no reason why a conditional proposition should not be true, though both the antecedent and the consequent are impossible; as if one were to say: *If a man is an ass, he has four feet*. Or, he may be understood to mean that God can do some things which now seem to be evil: which, however, if He did them, would then be good. Or he is, perhaps, speaking after the common manner of the pagans, who thought that men became gods, like Jupiter or Mercury.

⁹ This passage is taken from *The Basic Writings of St. Thomas Aquinas, op. cit.*, p. 264.

¹⁰ This passage is taken from S. N. Deane, *St. Anselm* (LaSalle, Open Court, 1958), p. 14.

In this passage St. Thomas offers three suggestions as to how the problem we are discussing might be solved. (I do not count the suggestion made in the last sentence of this passage because it is clear that St. Thomas is not here talking about *God* but about individuals such as Jupiter or Mercury who are mistakenly thought to be God by certain misguided pagans.) Let us look at these three suggestions:

(A) St. Thomas begins with the claim that "to sin is to fall short of a perfect action." He then says that an omnipotent being cannot fall short in action. The conclusion is that God cannot sin because He is omnipotent. Essentially this same reasoning is developed in slightly more detail in the seventh chapter of St. Anselm's *Proslogium*. Anselm says:¹⁰

But how art Thou omnipotent, if Thou are not capable of all things? or, if Thou canst not be corrupted and canst not lie . . . how art Thou capable of all things? Or else to be capable of these things is not power but impotence. For he who is capable of these things is capable of what is not for his good, and of what he ought not to do and the more capable of them he is, the more power have adversity and perversity against him; and the less has he himself against these.

Anselm concludes that since God is omnipotent, adversity and perversity have no power against Him and He is not capable of anything through impotence. Therefore, since God is omnipotent, He is not capable of performing morally reprehensible actions.

This argument is interesting. Both Thomas and Anselm agree that God is unable to sin. Their effort is to show that instead of being in conflict with the claim that God is omnipotent, the assignment of this inability is a direct consequence of this latter claim. However, I think that the reasoning fails. Let us agree that to the extent that an individual is such that "adversity and perversity" can prevail against him, to that extent is he weak—*morally* weak. He is then capable of "falling short in action," i.e., of doing "what he ought not to do." So far as I can see, an individual that is able to bring about any consistently describable state of affairs might well be morally weak. I can find no conceptual difficulty in the idea of a diabolical omnipotent being. Creative-power and moral strength are readily discernible concepts. If this is right, then it does not follow from the claim that God is omnipotent that He is unable to act in a morally reprehensible way. In fact, as was set out in the original statement of the problem, quite the opposite

conclusion seems to be warranted. If a being is able to bring about *any* consistently describable state of affairs, it would seem that he should be able to bring about states of affairs the production of which would be morally reprehensible. St. Thomas' first suggestion thus seems to be ineffective as a solution to the problem we are confronting. (I shall have something more to say on this topic in the fourth section of this paper.)

(B) The Philosopher says that God can deliberately do what is evil. Looking for a way of understanding this remark whereby it can be squared with his own view on the matter, St. Thomas suggests that what Aristotle may have meant is that the individual that is God can do evil *if He wants to*. Thomas adds that this last statement might be true even if it is impossible that God should want to do evil and even if it is also impossible that He can do evil. The point seems to be that although the statements "The individual that is God wants to do evil" and "The individual that is God can do evil" are false (or impossible), the conditional statement containing the first of these statements as the antecedent and the second of these statements as the consequent, might nonetheless be true.

Consider the statement: "Jones has an ace in his hand if he wants to play it." This statement has the surface grammar of a conditional, but it is not a conditional. The item mentioned in the "if . . ." clause does not condition the item described in the rest of the statement. If Jones has an ace in his hand, he has an ace in his hand whether or not he wants to play it. What, then, does the "if . . ." clause do in this statement? I think that it serves as a way of recording a certain indeterminacy as to what will be done about (or, with respect to) the unconditional fact described in the rest of the statement. Whether or not this last remark is precisely right, the major point to be seen here is this: The statement "Jones has an ace in his hand if he wants to play it" is false if the statement "Jones has an ace in his hand" is false. We are here dealing with a use of "if . . ." that does not fit the analysis usually given conditional statements such as "I shall be nourished if I eat."

Now consider the statement: "Jones can wiggle his ear if he wants to." I think that this is another instance in which "if . . ." operates in a non-conditional capacity. If Jones has the ability to wiggle his ear, he has the ability whether or not he wants to wiggle his ear. The question of whether he wants to wiggle his ear is independent of whether he has the ability to do so. As in the case

above, what the "if . . ." clause adds in this statement is not a condition on the claim that Jones has an ability. It serves as a way of recording the idea that there is some indeterminacy as to whether the ability that Jones has will be exercised. But again, I am less concerned with whether this last remark about the function of the "if . . ." clause is precisely right than I am with the relation between the truth values of "Jones can wiggle his ear if he wants to" and "Jones can wiggle his ear." In this case, as above, if Jones does not have the ability to wiggle his ear, then the statement "Jones can wiggle his ear if he wants to" is false. If the second of the above statements is false, then the first is false too.

St. Thomas says that the statement "God can sin if He wants to" is true. He adds that both the antecedent and the consequent of this conditional are "impossible." The trouble here, I think, is that "God can sin if He wants to" is not a conditional statement; and the most important point to be seen in this connection is that this statement is false if its component "God can sin" is false. But, St. Thomas clearly holds that "God can sin" is false (or impossible)—he says that God's inability to sin is a consequence of the fact that He is omnipotent. The conclusion must be that "God can sin if He wants to" is also false (or impossible). Thomas has not provided a way of understanding The Philosopher's claim that God can deliberately do what is evil. As long as St. Thomas insists that God does not have the ability to sin (which, he says, follows from the claim that God is omnipotent) he must deny that God can sin if He wants to. He must then reject The Philosopher's claim that God can deliberately do what is evil if this latter means that God can sin if He wants to.

(C) Still looking for a way of understanding the idea that God can deliberately do what is evil, St. Thomas' next suggestion is that God can do things which seem evil to us but which are such that if God did them, they would not be evil. I think that there are at least two ways of understanding this comment.

First, Thomas may be suggesting that God has the ability to bring about states of affairs that are, in fact, good, but which seem evil to us due to our limited knowledge, sympathy, moral insight, etc. However, even if we were to agree that this is true, Thomas could draw no conclusion as regards the starving-child situation described earlier. We have specified this situation in such a way that it not only *seems* evil to us, but *is* evil in fact. We have included in our description of this case that the child suffers

intensely; that this suffering is not deserved and that it might have been avoided. We have added that the suffering does not contribute to a greater good. Thus, this line of reasoning does not really help with the major problem we are discussing in this paper. We still have a range of consistently describable states of affairs that God (being perfectly good) cannot bring about. We thus still have reason to think that God (being perfectly good) is not omnipotent.

Secondly, St. Thomas may be suggesting that God has the ability to bring about *any* consistently describable states of affairs (including the starving-child situation), but that if *He* were to bring about such a situation, it would no longer count as evil. Let "evil" cover any situation which is such that if one were to (knowingly) bring it about (though it is avoidable), that individual would be morally reprehensible. The view we are now considering requires that we append a special theory about the meanings of the *other* value-terms involved in our discussion. In particular, it requires that when applied to God, the expressions "not morally reprehensible" and "perfectly good" be assigned meanings other than the ones they have when used to characterize individuals other than God. If a man were knowingly to bring about the starving-child situation, he would be morally reprehensible. He could no longer be described as perfectly good. But (so the argument goes) if God were to bring about the same situation, He might still count as perfectly good (not morally reprehensible) in the special senses of "not morally reprehensible" and "perfectly good" that apply *only* to God.

I have two comments to make about this second way of understanding St. Thomas' claim that God can do things that seem evil to us but which are such that if He did them, they would not be evil.

First, in my opinion the view we are now entertaining about the theological use of "perfectly good" and "not morally reprehensible" is one that was decisively criticized by Duns Scotus, Bishop Berkeley, and John Stuart Mill.¹¹ If God were to bring about circumstances such as the starving-child situation, He would be morally reprehensible and thus not perfectly good in the ordinary senses of these phrases. If we now contrive some special

phrases (retaining the tablettes "not morally reprehensible" and "perfectly good") that might apply to God though He produces the situation in question, this will be of no special interest. Whatever *else* can be said of God, if He were to bring about the starving-child situation, He would not be an appropriate object of the *praise* we ordinarily convey with the phrase "perfectly good." He would be an appropriate object of the *blame* we ordinarily convey with the phrase "morally reprehensible." We might put this point as follows: If we deny that God is perfectly good in the ordinary sense of "perfectly good," and if we cover this move by introducing a technical, well-removed, sense of "perfectly good" that can apply to God though He brings about circumstances such as the starving-child situation, it may appear that we have solved the problem under discussion in this paper, but we haven't. We have eliminated conflict by agreeing that God lacks one of the "perfections," i.e., one of the qualities the possession of which makes an individual better (more praiseworthy) than he would otherwise be. Unless a being is perfectly good in the *ordinary* sense of "perfectly good," that being is not as praiseworthy as he might otherwise be. It was the sense of "perfectly good" that connects with the idea of being morally praiseworthy (in the ordinary sense) that gave rise to the problem in the first place. Surely, it is this sense of "perfectly good" that religious people have in mind when they characterize God as perfectly good.

The second remark I should like to make about this second interpretation of St. Thomas' third suggestion is that the view assigned to St. Thomas in this interpretation is one that he would most likely reject. I shall need a moment to develop this point.¹²

Consider the word "triangle" as it occurs in the discourse of geometry. Compare it with "triangle" as it is used in the discourse of carpentry or wood-working. Within the discourse of geometry, the criteria governing the use of this term are more strict than are the criteria governing its use in the discourse of carpentry. The geometrical figure is an exemplary (i.e., perfect) version of the shape embodied in the triangular block of wood. We reach an understanding of the geometrical shape by correcting imperfections (i.e., irregularities) in

¹¹ See Scotus' *Oxford Commentary on the Sentences of Peter Lombard*, Q. II ("Man's Natural Knowledge of God"), second statement, argument IV; Berkeley's *Alciphron*, Dialogue IV, sects. 16-22 (especially sect. 17); and J. S. Mill's *An Examination of Sir William Hamilton's Philosophy*, ch. 6. What follows in this paragraph is what I think constitutes the center of these three discussions.

¹² The next two paragraphs are taken almost without change from the Introduction to my book *God and Timelessness* (London, Routledge and Kegan-Paul, 1969). What I say here about the relation between "triangle" as used in the discourse of geometry and "triangle" as used in the discourse of carpentry is very much like a thesis developed by John Stuart Mill in the text mentioned above.

the shape of the triangular block. Now, let's ask whether "triangle" has the same meaning in the two cases. We might answer this question in either way. Once the relation between the criteria governing its use in the two cases is made clear, no one would be confused if we were to say that "triangle" has the same meaning in the two cases, and on one would be confused if we were to say that "triangle" has different meaning in the two cases. Regarding the relation between the criteria, the following point seems to me to be of considerable importance: If a block of wood is triangular, it has three angles that add up (roughly) to 180 degrees and its sides are (roughly) straight. *At least this much* is implied with respect to a geometrical figure when one characterizes it as a triangle. By this I mean that if one could find reasons sufficient for rejecting the claim that a given thing is a triangle as "triangle" is used in the discourse of carpentry (suppose that one of its sides is visibly curved or suppose it has four angles), these same reasons would be sufficient for rejecting the claim that the thing in question is a triangle as "triangle" is used in the discourse of geometry. In fact, more than this can be said. If one could find slight irregularities in the shape of a thing that would cause some hesitation or prompt some reservation about whether it is a triangle as "triangle" is used in the discourse of carpentry, such irregularities would be sufficient to establish that the thing in question is not a triangle as "triangle" is used in the discourse of geometry.

According to St. Thomas, finite things are caused by God. They thus bear a "likeness" to God. God's attributes are exemplary-versions of the attributes possessed by finite things. We reach whatever understanding we have of God's attributes, by removing "imperfections" that attend these qualities when possessed by finite things.¹³ With respect to the predicate "good," St. Thomas writes as follows in the *Summa Theologica* (Pt. I, Q. 6, A. 4):¹⁴

Each being is called good because of the divine goodness, the first exemplar principle as well as the efficient and telic cause of all goodness. Yet it is nonetheless the

case that each being is called good because of a likeness of the divine goodness by which it is denominated.

Again, in *questiones disputatae de veritate* (XXI, 4), St. Thomas says:

Every agent is found to produce effects which resemble it. Hence, if the first goodness is the efficient cause of all things, it must imprint its likeness upon things which it produces. Thus each thing is called good because of an intrinsic goodness impressed upon it, and yet is further denominated good because of the first goodness which is the exemplar and efficient cause of all created goodness.

Shall we say that "good" has the same meaning when applied to God as it has when applied to things other than God (e.g. Socrates)? As above, it seems to me that the answer we give to this question is unimportant once we get this far into the discussion. We might say that "good" has the same meaning in the two cases, and we might say that it has different meanings in the two cases. We might even say (as St. Thomas sometimes says) that we are here dealing with a case in which "good" is "midway between" having the same meaning and having different meanings in the two cases. However, as above, the following point has importance regardless of how one answers the question about same or different meanings. When St. Thomas affirms that God is good, I think he means to be saying *at least as much* about God as one would say about, e.g., Socrates, if one were to affirm that Socrates is good. A study of "good" in non-theological contexts reveals at least the minimum implications of the corresponding predication statements relating to God. If we could find reasons sufficient for rejecting the claim that a given thing is good as "good" is used in discourse about finite agents, these same reasons would be sufficient for rejecting the claim that the thing in question is good as "good" is used in discourse about the nature of God. In fact, if we could find moral irregularities sufficient to cause hesitations or prompt reservations about whether a thing is good as "good" is used in discourse about finite agents,

¹³ In the *Summa Theologica* (Pt. I, Q. 14, a.1) St. Thomas says:

Because perfections flowing from God to creatures exist in a higher state in God Himself, whenever a name taken from any created perfection is attributed to God, there must be separated from its signification anything that belongs to the imperfect mode proper to creatures. (Quoted from *The Basic Writings of St. Thomas Aquinas, op. cit.*, p. 136.)

¹⁴ Both of the following passages were translated by George P. Klubertanz, S. J., *Thomas Aquinas on Analogy* (Chicago, Loyola Press, 1960), p. 55. For a good analysis of St. Thomas' views on the topic of theological predication, see the whole of Klubertanz' discussion in ch. III. For an enlightening discussion of how poorly St. Thomas has been understood on this topic (even by his most illustrious interpreters) see Klubertanz' remarks in ch. I and Berkeley's discussion of St. Thomas in *Alcephron*, IV, sects. 20-22. According to Berkeley, St. Thomas' doctrine of "analogy by proportionality" is to be regarded as an expression of the view we are now discussing.

these irregularities would be sufficient to establish that the agent under consideration is not good as "good" is used in the discourse of theology.

So far as I can see, St. Thomas would not endorse a technical, well-removed sense of the phrase "perfectly good" that could apply to God even if God were to bring about circumstances or states of affairs the production of which would be morally reprehensible (in the ordinary sense of "morally reprehensible"). When St. Thomas says that God is good, he means to be saying that God possesses the exemplary version of the quality assigned to Socrates in the sentence "Socrates is good." This is to say that while Socrates is good, God is *perfectly* good. But on this understanding of the matter, God could not be perfectly good were He to bring about the starving-child situation described earlier. If Socrates were to bring about such a situation, we would probably refuse to describe him as "good." At the very least, we would surely have hesitations or reservations concerning his moral goodness. But, if such an action would be sufficient to cause hesitations concerning an application of "good" in discourse about finite agents, this same action would be sufficient to *defeat* an application of "good" in discourse about the nature of God. In this latter context, "good" means "*perfectly* good." The logic of this phrase will not tolerate even a minor moral irregularity.

III

I want now to discuss an approach to our problem that is very different from any of those suggested by St. Thomas. It is an approach taken by J. A. McHugh in the *Encyclopedia* article mentioned above. I think we can best get at the center of McHugh's thinking if we start with a review of that side of the problem generated by the concept of perfect goodness.

God is perfectly good. This is a necessary statement. If a being is perfectly good, that being does not bring about objects or states of affairs the production of which would be morally reprehensible. This, too, is a necessary truth. Thus, the statement "God brings about objects or states of affairs the production of which would be morally reprehensible" is logically contradictory. It follows that God cannot bring about such states of affairs. But, McHugh argues, this should not be taken as a reason for denying God's omnipotence. As St. Thomas has pointed out, a being may be omnipotent and yet not be able to do an act whose

description is logically contradictory. (A being may be omnipotent though he is not able to make a round-square.) Since the claim that God acts in a morally reprehensible way is logically contradictory, God's inability to perform such acts does not constitute a limitation of power.

Consider the following argument: The term "Gid" is the title held by the most efficient of those who make only leather sandals. "Gid makes leather belts" is thus a logical contradiction. It follows that the individual that is Gid cannot make leather belts. But Gid may still be omnipotent. Though He does not have the ability to make leather belts, our analysis of "omnipotence" requires only that an omnipotent being be able to do an act whose description is logically consistent and "Gid makes leather belts" is logically inconsistent. Thus, Gid's inability to make leather belts does not constitute a limitation on his power.

I think it is plain that this last argument is deficient since its conclusion is absurd. I think, too, that in this case, two difficulties are forced pretty close to the surface.

First, the description of the kind of object that Gid is (allegedly) unable to make (viz., leather belts) is not logically contradictory. What is contradictory is the claim that *Gid makes them*. But our definition of "omnipotent" requires only that the *state of affairs* brought about be consistently describable (excluding, therefore, round squares). It does not require that a statement in which it is claimed that a given individual brings it about be consistent. Thus, if Gid does not have the ability to produce leather belts, he is not omnipotent on St. Thomas' definition of "omnipotent." If it follows from the definition of "Gid" that the individual who bears this title cannot make leather belts; and if this entails that the individual in question does not have the creative-ability to make belts, the conclusion must be that, by definition, the individual who bears this title is a limited being. I think the same kind of conclusion must be drawn in the case of God's ability to sin. If it follows from the definition of "God" that the individual bearing this title cannot bring about objects or states of affairs the production of which would be morally reprehensible; and if it follows from this that the individual bearing this title does not have the creative power necessary to bring about such states of affairs though they are consistently desirable; the conclusion is that the individual who is God is not omnipotent on the analysis of "omnipotent" that we are supposing. The fact (if it is a fact) that this

creative limitation is built into the definition of "God" making "God sins" a logical contradiction does not disturb this conclusion. The upshot is, simply, that the term "God" has been so specified that an individual qualifying for this title could not be omnipotent. (Of course, this is awkward because it is also a condition of bearing this title that the individual in question be omnipotent.)

The second difficulty in the argument about *Gid* is this: The term "*Gid*" has been defined in such a way that "*Gid* makes leather belts" is logically contradictory. The conclusion drawn is that *Gid* *cannot* make belts. What this means is that if some individual makes leather belts, this is logically sufficient to assure that the individual in question does not bear the title "*Gid*." But it does not follow from this (as is supposed in the argument) that the individual who is *Gid* does not have the *ability* to make leather belts. All we can conclude is that if he does have this ability, it is one that he does not *exercise*. Thus, as is affirmed in the argument, the individual who is *Gid* might be omnipotent though he cannot make leather belts (and be *Gid*). If we suppose that he is omnipotent, we must conclude that he has the ability to make belts; but since, by hypothesis, the individual in question is *Gid*, we know (analytically) that he does not exercise this ability. Again, I think the same is true with respect to the argument about God's inability to sin. The term "God" has been so specified that the individual who is God *cannot* sin and be God. But it will not follow from this that the individual who is God does not have the *ability* to sin. He might have the creative power necessary to bring about states of affairs the production of which would be morally reprehensible. He is perfectly good (and thus God) insofar as He does not exercise this power.

IV

If we collect together a number of threads developed in the preceding discussions, I think we shall have enough to provide at least a tentative solution to the problem we have been discussing. I shall proceed by distinguishing three ways in which the statement "God cannot sin" might be understood.

"God cannot sin" might mean: "If a given individual sins, it follows logically that the individual does not bear the title 'God'." In this case, the "cannot" in "cannot sin" expresses logical impossibility. The sentence as a whole might be rewritten as follows: $N(x)$ (If x is God then x does

not sin.) On the assumptions we are making in this paper, this statement is true. We have supposed that the meaning of the title term "God" is such that it is a logically necessary condition of bearing this title that one be perfectly good and thus that one not perform actions that are morally reprehensible.

Secondly, "God cannot sin" might mean that if a given individual is God, that individual does not have the ability to sin, i.e., He does not have the creative power necessary to bring about states of affairs the production of which would be morally reprehensible, such as, e.g., the starving-child situation described earlier. In this case, the "cannot" in "cannot sin" does not express logical impossibility. It expresses a material concept—that of a limitation of creative-power (as in, e.g., "I cannot make leather sandals"). On St. Thomas' analysis of "omnipotence" if the individual who is God (Yahweh) cannot sin in this sense, He is not omnipotent. Further, I think there is strong reason to suspect that if the individual that is God (Yahweh) cannot sin in this sense, He is not perfectly good either. Insofar as the phrase "perfectly good" applies to the individual that is God (Yahweh) as an expression of praise—warranted by the fact that this individual does not sin—God could not be perfectly good if He does not have the ability to sin. If an individual does not have the creative-power necessary to bring about evil states of affairs, he cannot be praised (morally) for failing to bring them about. Insofar as I do not have the physical strength necessary to crush my next door neighbor with my bare hands, it is not to my credit (morally) that I do not perform this heinous act.

Thirdly, "God cannot sin" might mean that although the individual that is God (Yahweh) has the ability (i.e., the creative power necessary) to bring about states of affairs the production of which would be morally reprehensible, His nature or character is such as to provide material assurance that He will not act in this way. This is the sense in which one might say that Jones, having been reared to regard animals as sensitive and precious friends, just *cannot* be cruel to animals. Here "cannot" is not to be analyzed in terms of the notion of logical impossibility and it does not mark a limitation on Jones's physical power (he may be physically able to kick the kitten). It is used to express the idea that Jones is *strongly disposed* to be kind to animals, or at least to avoid actions that would be cruel. We have a special locution in English that covers this idea. When we say that Jones cannot be cruel

to animals, what we mean is that Jones cannot *bring himself* to be cruel to animals. On this third analysis of "God cannot sin," the claim conveyed in this form of words is that the individual that is God (Yahweh) is of such character that he cannot bring himself to act in a morally reprehensible way. God is strongly disposed to perform only morally acceptable actions.

Look back for a moment over the ground we have covered.

McHugh noticed that the statement "God sins" is logically incoherent. He thus (rightly) concluded that God cannot sin. He was here affirming that the semantical import of the title term "God" is such that an individual could not (logically) bear this title and be a sinner. McHugh's conclusion ("God cannot sin") was thus intended in the first sense just mentioned. But McHugh then went on to suppose that God cannot sin in a sense of this phrase that connects with the notion of omnipotence. This is the second sense mentioned above. This conclusion was not warranted. The individual who bears the title "God" (Yahweh) might have the creative power necessary to bring about objects or states of affairs the production of which would be morally reprehensible even though "God sins" is logically contradictory. The conclusion is, simply, that if an individual bears the title "God," He does not exercise this creative-power.

St. Thomas and St. Anselm said that God cannot sin in that "adversity and perversity cannot prevail against Him." This appears to be the claim put forward in the *Epistle of James* 1:13—the claim embodied in the theological doctrine of God's impeccability—viz., "God cannot be tempted by evil." The individual that is God has a very special kind of strength—moral strength, or strength of character. He is, as we say, "above temptation." Both Thomas and Anselm concluded that God's

inability to sin has a direct connection with the notion of omnipotence. It is because God is omnipotent that He is unable to sin. This line of reasoning confuses the second and third senses of the statement "God cannot sin." If we say that the individual who is God cannot sin in this second sense (i.e., in the sense that connects with the idea of creative power and thus with the standard notion of omnipotence) this is not to assign that individual strength. It is to assign Him a very definite limitation. The strength-concept in this cluster of ideas is the notion of not being able to *bring oneself* to sin. God has a special strength of character. But this latter concept is expressed in the third sense of "God cannot sin." As I argued earlier, this third sense appears to have no logical connection with the idea of having or lacking the creative power to bring about consistently describable states of affairs. It thus appears to have no logical connection to the notion of omnipotence as this latter concept is explicated by St. Thomas.

The individual that is God cannot sin and bear the title "God." The individual that is God cannot sin in that sinning would be contrary to a firm and stable feature of His nature. These claims are compatible with the idea that the individual that is God has the ability (i.e., the creative power necessary) to bring about states of affairs the production of which would be morally reprehensible. All we need add is that there is complete assurance that He will not exercise this ability and that if He did exercise this ability (which is logically possible but materially excluded), He would not bear the title "God." Further, if God is to be omnipotent in St. Thomas' sense of "omnipotent," and if God is to be perfectly good in a sense of this phrase that expresses praise for the fact that He refrains from sinful actions, this appears to be the conclusion that *must* be drawn.

V. STATEMENTS AND PERFORMATIVES

JEREMY D. B. WALKER

AUSTIN'S lectures on *How to Do Things with Words* (Oxford, 1962) were partly aimed at destroying the assumption of philosophers that the business of a "statement" can only be to "describe" some state of affairs, or to "state some fact," which it must "do either truly or falsely" (p. 1). He attacks this "age-old assumption" twice: first in the course of his long account of performative utterances, especially on pp. 45-52 and 77-91; and again in Lecture XI. Here he uses the notions constructed in the second part of his lectures out of the ruins of the earlier account.

Austin begins by describing quite informally the traditional assumptions he wants to destroy, and introducing his own concept of the performative utterance. He then tries first to sharpen and next to blur some putative distinctions between them. One effect of this blurring is to show, or suggest very strongly, that the special theory of performative utterances should be abandoned in favor of a more general theory of speech acts. Some kind of *Aufhebung*, though it is unclear exactly what, is called for.

In this paper I shall argue that Austin's attempts to blur his initial distinction between the performative and the "constative" utterances fail. So, as far as this part of his work goes, we need repudiate neither "statements" and their "truth-values," nor the concept of a performative utterance. (This conclusion need not significantly affect our assessment of the theory of illocutionary acts.)

Austin's introduction of his theory of performatives is brief and not very clear. For he does not give any clear description of those notions his theory is intended to account for, e.g., statement, description, fact, or truth. Still, it is plain that he meant his theory to be added to what he regarded as a number of attacks on the "constative fallacy." This fallacy has several aspects. But its general form is something like this: a philosophical assumption that all or most forms of words of a certain kind are "statements of fact." Performative utterances are characterized, on their introduction (pp. 4 ff.), as "utterances which can fall into no hitherto recognized *grammatical* category save that

of 'statement' . . . yet such that (A) they do not 'describe' or 'report' or constate anything at all, are not 'true or false'; and (B) the uttering of the sentence is, or is part of, the doing of an action, which again would not *normally* be described as saying something." But Austin offers no arguments that his paradigm cases (marrying, naming, bequeathing, betting) fit these descriptions.

Austin opposes to the traditional "truth or falsity" of the statement the new concepts of the "happiness" and "unhappiness" of the performative utterance. But he then proceeds to describe several *similarities* between important features of the performative and features normally ascribed to statements. Now these were criteria for distinguishing performatives from "constatives." So the greater the similarities, the less use will such features be as distinguishing criteria and the less clear and sharp will the original distinction be. This is exactly the point he intends to make.

The Discussion on pp. 45-55

Austin begins by explaining that his doctrine of "infelicities," i.e., of the conditions which have to be satisfied for a performative to be "happy," is equivalent to a doctrine that "for a certain performative utterance to be happy, certain statements have to be true" (p. 45). For instance, if in saying "I apologize" I actually do apologize, at least four things must be true: (i) that I am doing something including apologizing, (ii) that certain conditions mentioned in the A-type rules obtain, (iii) that certain Gamma-type conditions obtain, e.g., that I am truly penitent, and (iv) that I am committed to doing something subsequently. But the doctrine is also equivalent to a doctrine about "the sense in which 'I apologize' implies the truth of each one of these" (p. 46). So the description of performatives can be taken as a description of what such utterances *imply*, and how they do it. And this allows us to compare the "implications" of performatives with the "implications" of statements. Conversely, to claim that statements have "implications" is to claim that they have logical features *similar* to those logical features of performatives by

which they are capable of being "happy or unhappy." What are the alleged similarities?

Austin mentions three kinds of statement "implication," viz., those logical relations ordinarily called entailment, implication (in Moore's sense), and presupposition. Entailment is, roughly, a matter of one proposition's truth being inconsistent with another's. Implication is a relation between an assertion and a belief and not concerned with the consistency of propositions. The idea of presupposition is illustrated by the claim that "All Jack's children are bald" presupposes that Jack has some children. To deny a relation of presupposition is to "outrage speech" in some way. (These views are fully set out on pp. 47-50.)

Even before comparing these kinds of implication with their alleged counterparts for performatives, we should notice that the types of error indicated by Austin are both more numerous than those catered for by the traditional account, and importantly different in that they involve not just the words uttered, but also the uses of these words, the acts of making these statements. This is just the sort of consideration which leads Austin up to his eventual attempt to construe statement-making as simply one kind of speech-act.

He now proceeds to argue that the second, third, and fourth types of "implication" which performatives are said to have are similar, respectively, to the relations of presupposition, implication, and entailment.

Presupposition

When the "presupposition" of a statement is false, Austin claims, we nowadays tend to say that the statement is not itself false but not meaningless either. "Here I shall say 'the utterance is void'," (see pp. 50 ff.). He goes on: "Compare this with our infelicity when we say 'I name . . .,' but some of the conditions (A.1) and (A.2) are not satisfied. . . . Here we might have used the 'presuppose' formula: we might say that the formula 'I do' presupposes lots of things: if these are not satisfied the formula is unhappy, void. . . ."

Compare two utterances, (1) "The King of France is bald," and (2) "I bequeath my goods to the King of France." (I am choosing cases where the causes of the breakdowns are as alike as possible.) What shall we say about (2)? Despite use of the correct formula, the utterance cannot be counted as performance of the act of bequeathing one's property. In this sense it may be called "void." But the question of the truth or falsity of the speaker's

words, never having arisen in the first place, cannot be affected by the breakdown. Contrast this with the "parallel" results for (1). Here the truth or falsity of the utterance, i.e., of the words uttered, is affected. The speaker's performance cannot be counted as an act of *saying something true or false*. This does not mean, as Austin seems to imply, that it cannot be counted any longer as a statement. It is all very well to use the term "void" for this case too; but this simply avoids the question, void in its purported capacity as *what sort of performance?* Austin seems to assume that an utterance cannot be called a statement unless it is either true or false, and I see no justification for this assumption.

He claims, for instance, that we cannot say "The King of France is bald, but there is no King of France." But we can *say* this. There may be, as he says, something outrageous in saying it, but unless we could say it the possibility of the outrage could not arise. And normally someone who says this will be held to have *stated* that the King of France, etc. Now it is true that *what* someone who says this says cannot be either true or false, at least on the current view Austin expresses. But if it makes sense to claim that we cannot say "I bequeath my goods to the King of France, but there is no King of France," it is a very different sense. We can obviously again *say* this, but if the second part of the utterance is true, the first part cannot count as the *performance* it specifies. (Note that the utterance has to be taken as two distinct parts, since the first clause cannot be either true or false, and the second cannot be construed as a performative—except in the sense, if any, in which all statements are performatives.) So I conclude that Austin has not yet shown that statements are not unique in their capacity for truth and falsity, nor that performatives are not unique in the capacity for happiness and unhappiness of the kind indicated.

Implication

Austin's paradigm is Moore's "*p*, but I don't believe that *p*." He claims that this is clearly a case of insincerity, and that the "unhappiness" is just the same as the unhappiness of saying "I promise" when I do not intend to do what I promise. "The insincerity of an assertion is the same as the insincerity of a promise" (p. 50).

Compare two utterances, (1) "The cat is on the mat" (said disbelieving it), and (2) "I promise I'll come" (said not intending to). Now in his Lecture 1 Austin claimed that in (2) the promise is not void, nor a lie or a misstatement, although it is

perhaps misleading. So he saw the difference made here by the statement's capacity for truth and falsity. Philosophy students are always taught that promises, as contrasted with statements, cannot be true or false, and he began by endorsing this view. But just this is fatally blurred by his later attempt to subsume both cases under the heading of "insincerity." (Nor is it even true that the statement kind of case is always "insincere," by any means.) And calling *both* kinds of unhappiness further blurs the normal distinction. For, although it is odd to call insincere promises "unhappy" *as promises*, it is arguable that the insincerity does cast a doubt upon the actual promising, i.e., allow the question, did he really promise or not? But there is nothing "unhappy" in this way about insincere statements. The speaker's insincerity casts no doubt at all on whether he really said (stated) what he did. This is just the real unhappiness of the situation!

Even if both statements and promises imply something about the speaker's mind in the same sense of "imply," still there is an important difference in *what* they imply. Saying "the cat is on the mat" implies not just that I believe this: it implies that I believe what I say is *true*—a concept peculiarly bound up with belief. And there is no feature analogous to this with promising.

And what is the point in Austin's claim that saying '*p*' together with "I don't believe that *p*" is "not possible" (p. 49), and how is this parallel to a feature of promising? It is certainly possible to *say* both of these things, and even to say both truly, since their truth is not incompatible. Might it be that we cannot *assert* both together? But this is too vague, since the concept of assertion is not at all clear. Might it be that we cannot express two beliefs together in these words, since they cannot be understood as representing two compatible beliefs? Or that both of these statements could not be made sincerely at the same time? Unhappily, these two formulations are not obviously true either. But anyway, whatever precisely the facts about statements, they contrast quite clearly with the facts about promising. For one thing, although it is possible to say "I promise" and "but I don't intend to" together, it is not possible to *say both* false, since "I promise" can be neither true nor *truly*. Might the point be, then, that we cannot at the same time *say* the second and *do* what is specified in the first? But this would conflict with Austin's view that the utterance of the words "I promise" *constitutes*, in the right conditions, the act of promising. Of course, if the second utterance is

true, the first cannot be sincere. But this sounds like a disguised tautology. So I conclude that here too Austin has not identified anything that blurs the sharp distinction between statements and performatives.

Entailment

Here Austin considers the way in which "I promise" implies "I ought." He is careful to say that this is not *the same* as entailment, but claims that it is "parallel." "'I promise but I ought not' is parallel to 'it is and it is not' . . . Just as the purpose of assertion is defeated by an internal contradiction . . . the purpose of a contract is defeated if we say 'I promise and I ought not' . . . It is a self-stultifying procedure. One assertion commits us to another assertion, one performance to another performance" (p. 51). What is this parallel?

Now the idea of entailment is by no means itself clear, nor is Austin's explanation of it. To claim that we cannot *say* "All men blush but not any men blush" is not helpful. For we certainly can say this, and it is what precisely is wrong with saying this, i.e., wrong with *what* is said, that we want explained. Austin's glosses here are not much use. First he tells us that we cannot say this because the first clause is the *contradictory* of the second. But this is surely to use a different word in which the same problem hides. Next, he tells us that entailment is really a relation between the *truth* of the first proposition and the truth of the second. So presumably we cannot say "All men blush, etc.," because the truth of the first clause is inconsistent with the truth of the second. This is not much more help, except that it does rightly bring out that the concept of entailment involves in its elucidation the concepts of truth and falsity.

Later Austin adds that the purpose of assertion is defeated by an internal contradiction (p. 51). This refers to that connection between "contradiction" and truth which still needs clarifying. But can he simply mean that one (logically) cannot assert a self-contradictory statement? As before, this leads us into murky waters. Still, the important point is that what "cannot be said" (asserted, believed) depends on the truth-values of what is said, and not the other way round.

He says, again, that the assertion "All men blush" *commits* the speaker to the further assertion "Some men blush." But it "commits" him only in the sense that, if what is first asserted is true, then to assert the second would also be to assert something true. The speaker is not "committed" to

making this second assertion in any normal sense of the word.

With the performative "I promise" things are very different. For the idea of the truth-value of what the speaker says plays no part in determining what he is committed to, even if it applies at all. Again, here it is the act of *saying* "I promise" that commits the speaker to the action mentioned. But in the case of statement-entailment, it is not *that* one utters the proposition but *what* one utters, the proposition itself, that entails some further proposition. Again, Austin claims a parallel between (i) saying "I promise" and not doing it, and (ii) saying '*p*' and saying 'not-*p*.' In case (i) it is not *doing* what one mentioned in the promise that is "inconsistent" with carrying out that commitment. "He carried out his commitment" and "He did not do what he promised" are, perhaps, logically inconsistent descriptions of his conduct. They are statements both of which cannot be true. But there is no sense in which *saying* "not-*p*" is inconsistent with saying '*p*.'

Consider again Austin's claim that saying "I promise . . . and (but) I am under no obligation to" defeats the purpose of the contract. One answer is: even if you follow a promise by saying "But I don't consider myself bound by that," you are still nonetheless bound by your promise. But anyway this merely shows another great difference between performatives and statements. For Austin's claim is plausible when he refers to the purpose of a contract or contradicting: namely, to give the second party a reason to believe you will act in a specified way and a foothold for penalizing you if you fail so to act. But there is no purpose, in this sense, to making statements, or asserting or "saying" things. There are many different possible purposes. I suppose one might claim (implausibly) that primary is the purpose of expressing your beliefs, or communicating what you believe to be true. And this seems to underlie Austin's obscure views about assertion.

Conclusion

Austin has not yet indicated anything about statements or performatives which blurs their original distinction. We may still ascribe truth-values to statements but not to performatives. And much of the plausibility of his claim that they have *similar* logical features rests on his tendency to blur the distinction between the *act* of uttering a statement and *what* is so uttered—what used to be called

a proposition. Truth and falsity can belong only to what is uttered, not to the uttering of it: happiness and unhappiness to the (performative) uttering, not to what is uttered. So far there is no good reason to abandon these useful distinctions.

The Discussion on p. 46ff and pp. 54ff

Here Austin discusses the first type of "implication"-relation he ascribes to performatives. He compares it to that relation to the facts in virtue of which statements are true or false. Now at first sight there seems to be a clear contrast. Statements, roughly speaking, depend for their truth-value on what is the case: performatives, by being happy or unhappy, *make* something the case. "John is running" depends for its truth on the fact that John is running: the fact that I am apologizing depends on the happiness of the utterance "I apologize."

Austin, however, tries to argue that this contrast is eroded from two sides. (1) "Connected with the utterance (constative) 'John is running' is the statement 'I am stating that John is running': and this may depend for its truth on the happiness of 'John is running,' just as the truth of 'I am apologizing' depends on the happiness of 'I apologize'" (p. 55). And (2), "connected with the performative . . . 'I warn you that the bull is about to charge' is the fact, if it is one, that the bull is about to charge: if the bull is *not*, then indeed the utterance . . . is open to criticism—but not in any of the ways we have hitherto characterized as varieties of unhappiness . . . we should feel much more inclined to say the warning was false or (better) mistaken, as with a statement" (p. 55). What is the force of these views?

Let us, for clarity, list some propositions Austin wants to assert:

- (i) If the performative "I apologize" is *happy*, then (it is a fact that) I am apologizing.
- (ia) If the performative "I apologize" is *happy*, then (the statement?) "I am apologizing" is *true*.
- (ii) If (it is a fact that) John is running, then (the statement) "John is running" is *true*.
- (iia) If (the statement) "John is running" is *happy*, then (the statement) "I am stating that John is running" is *true*.
- (iii) If the performative "I warn you that the bull is about to charge" is *happy*, then I am warning you, etc., and the statement "I am warning you, etc." is *true*.

- (iia) If the bull is about to charge, then the performative "I warn you, etc.," is (true or) not mistaken. If it is not, the performative is (false or) mistaken.

Austin argues that, although at first we want to distinguish sharply the phenomena exhibited by propositions (i) and (ia), on reflection we shall see that the distinction cannot be made sharply. For the phenomena exhibited by (ia) are similar to those exhibited by (iia), which is closely connected with (ii). And the phenomena exhibited by (iia) are similar to those exhibited by (ii) and the corresponding negative proposition. On the happiness of a performative depends the truth of some statement, i.e., the statement that the speaker is performing the act achieved in a happy utterance of that performative. On the "happiness" of the constative depends the truth of a further statement, i.e., the statement that the speaker is performing the act achieved in a happy utterance of that constative, i.e., stating that On the existence or non-existence of a fact depends the truth of a statement. On the existence or non-existence of a fact depends whether a performative is mistaken or not.

Unhappily, Austin does not explain the crucial notions in this account, i.e., those notions whose use enables him to construct the apparent parallels his argument rests on. He does not explain what calling a statement "happy" or "unhappy" can mean, and so fails to give sense to proposition (iia). He does not explain the connection between the statements "John is running" (say) and "I am stating that John is running"—surely an odd form of words! And *pro tanto* the sense of (iia) is still further obscured. He does not explain how a statement of the form "I am stating that *p*" has a truth-value, nor on what facts this depends, nor how it is connected with the truth-value of '*p*.' He does not explain the sense of the statement "I am apologizing"—again an odd form of words. And its sense cannot be read off from an understanding of the sense of "He is apologizing." Finally, he does not explain or justify his referring to the performative "I warn you that *p*" as true or false, mistaken or not mistaken, depending on *p*'s truth-value.

For the performative, i.e., the *act* of saying "I warn you that *p*" cannot be true or false. These are the wrong sort of predicates for characterizing *acts*, even speech acts. They characterize *what* is said in a speech act: e.g., not the act of stating '*p*,' but what is so stated. Perhaps this is why Austin is chary of saying that "I warn you, etc.," is "true or

false." On the other hand, "mistaken" and "not mistaken" characterize the *speaker*, not his speech act nor what he says. To be "mistaken" in a constative utterance is, roughly, to say something false while believing it to be true (or *vice versa*). But this will not do as it stands for performatives. If I say "I warn you, etc.," and the bull is not about to charge, it is not that I have said something false. But what I said certainly *implies* that the bull was about to charge. And if this is not the case, then this implied belief is false, and I can be called "mistaken," though not mistaken in what I actually say.

It is significant and important that Austin here turns from the clearly performative "I apologize" to the less clearly performative "I warn you that." For the *that*-clause imports the possibility of *something* being true or false. "I apologize" has no feature on which these notions can get a grip. Austin has tacitly singled out a *special class of performatives*, distinguished by this connection with the notions "true" and "false." Their similarity (if any) to constatives will not necessarily belong to other kinds of performative.

Now a good many of the verbs carrying illocutionary force that Austin lists in Lecture XII either normally do or can take the *that*-construction. For instance: find that, hold that, and rule that; warn that, urge that, and advise that; promise that, vow that, and bet that; affirm that, report that, and deduce that; and so on. This class of performatives is not small.

Further, very many of these verbs seem to belong to the class of "parenthetical" verbs in a certain respect. Instead of saying "I warn you that the bull is about to charge," we can equally well say "The bull is about to charge, I warn you." Instead of "I promise you that I'll come tomorrow," we can equally well say "I'll come tomorrow, I promise you." "Equally well," since the *force* of each of these pairs seems to be the same as the force of its twin.

Such "parenthetical" utterances clearly fall into two parts: a *statement* capable of truth or falsity, and an explicit performative formula. These are logically idiosyncratic forms. In particular, they are not *truth-functions* of the statements they contain, since *as a whole* they cannot be either true or false. Even though something is asserted *in* such an utterance, the utterance as a whole is not *asserted*. Now all of this throws light on the relation between explicit and implicit performatives. For the utterances "*p*, I warn you" and '*p*!' are clearly

closely connected, as Austin points out. The *statement that p* can be made with the *force* of a warning, i.e., uttered as a warning. As such it does not cease to be true or false. But it does not follow that the explicit addition of "I warn you (that)" leaves this logical feature unchanged. Even if '*p*!', uttered with the force of a warning, is a statement, it does not follow, and is false, that "*p*, I warn you" is a *statement*, any more than "I warn you that *p*" is a *statement*.

So Austin's argument that some kinds of performative are logically analogous to statements is not surprising. For, I have argued, their *performative force* can be expressed implicitly in the word-form of a *statement*. But this does not blur the distinction between forms of words which can be true or false and forms of words which can be merely happy or unhappy, e.g., "I warn you that *p*." And we can also now see why Austin claimed without argument that statements too can be happy or unhappy. For the form of words '*p*', e.g., "The bull is about to charge," certainly counts as a *statement* by most tests, but it can be given the force of a warning. Only we must distinguish still between the *act* of uttering these words with the force of a warning, and *what* is so uttered, the words. It is the *words* that are "true or false," the *act* that is (plausibly) "happy or unhappy." And even if this account is accepted, it is only the act of uttering these words *with a performative force* that is "happy or unhappy." It has not been shown that the act of uttering *the same words as a statement* can be either happy or unhappy.

So I conclude that Austin has not yet shown that happiness-unhappiness considerations affect statements, or that truth-falsity considerations affect performatives. His conclusion that the constative-performative distinction breaks down need not be accepted. The most that might be said is that his arguments imply a distinction not clearly made at the beginning of his Lectures. The distinction between the act of saying something and what is so said must be taken to cover a distinction between the act of making a statement and what is said in making a statement. (This is a point Austin actually goes on to make.)

The Discussion on pp. 59-90

Now Austin considers whether performatives might be distinguished by some criterion of grammar or vocabulary. He argues that there is no one such criterion, although there may be a complex

criterion. More important, however, are considerations underlying certain grammatical features of performatives, particularly the "initial favouritism for verbs in the so-called 'present indicative active'" (p. 60). Underlying this feature is the obvious idea that "actions can only be performed by persons, and obviously in our cases the utterer must be the performer . . ." (p. 60). Austin then argues that performatives ought to be "reducible, or expandible, or analyzable" into a form with a verb in the first person singular present indicative active. And that this verb will have a peculiar and special use, notably in that we shall find a systematic asymmetry between first person singular present indicative uses and uses in other persons or tenses.

But this new attempt has some difficulties. First, it appears to allow the utterance "I state that *p*" to be a *performative*. Later Austin accepts this (see p. 90 and p. 133). But his reasons are not clear. It is just because to say "I state that" is to state that? Or is it because to state that . . . is in some sense clearly to perform an *action* of a specific sort? Whatever his reasons, we should note that to classify stating as a performative is to withdraw one of the original distinguishing criteria of the performative: namely, its not "describing" or "reporting" or constating (see p. 5). Stating certainly fails this "semantic" criterion. And it is only because Austin has concluded, on the previous pages, that the latter breaks down that he can begin to consider stating as a possible performative at all. If I am right that those arguments were insufficient, then he did not show the semantic criterion to be inadequate, and thus is not justified in here considering stating as a performative. It is already ruled out.

Worse, in giving up the sharp distinction between statements, as true or false, and performatives, as happy or unhappy, Austin is tacitly giving up one of the only two explicit *general* criteria for the performative. The new "quasi-grammatical" criterion is quite different from either. And he gives no grounds for identifying the two concepts of a performative picked out (a) by the original two criteria and (b) by the second of those plus the new quasi-grammatical test. If we call (a) "type 1 performatives," and (b) "type 2 performatives," then even if stating counts as a type 2 performative it does not necessarily (and does not) count as a type 1 performative: though its being excluded from type 1 performatives does not entail its also not being a type 2 performative.

However, one might well doubt whether "I

state that" is a form of words we ever normally use. (We do use the forms "He stated that" and "Are you prepared to state that?" and even "May I state that.") So I doubt whether the verb "state" does clearly fall under even the new criterion. And if it does not, then stating will not even count as a type 2 performative.

A much greater difficulty with this new attempt is that often the same sentence can be used on different occasions as a performative and a constative. This shows that "no simple grammatical mark of utterances as *they stand*" (p. 67) can suffice to distinguish performatives from constatives. But worse, Austin notices certain cases where one and the same *formula* can be sometimes an explicit performative and sometimes "descriptive." He cites "I approve" as an example. He calls this phenomenon "a shift from descriptive to performative utterance and wavering between them" (p. 85). And he claims that it is found in three kinds of performatives, namely *behabitives*, *expositives*, and *verdictives*.

He says that there are three more or less distinct classes of *behabitives*. (1) Verbs like "thank" and "apologize," which are normally used in an explicit performative formula. (2) Verbs like "am grateful," which are "not pure but half descriptive" (p. 79). And (3) verbs like "feel grateful," which are merely descriptive. And he seems to distinguish three similar classes of *expositives* too. What does this show?

Obviously the main difficulty arises over the use of one and the same verb, e.g., "approve" or "agree," in all three kinds of construction. But just what is the difficulty? For we still have the two original criteria for the performative, or at least the second of them. Can we not still ask about *any* use of such a verb, (i) whether the utterance as a whole can be true or false, and (ii) whether issuing the utterance can be counted as "doing" something? Now Austin does propose here four informal tests for the genuine (pure) performative and the genuine (pure) constative. And they implicitly rely on those original criteria. For corresponding to the first criterion is the fourth of these tests, namely "to ask whether what one says could be literally false . . . , or could only involve insincerity (unhappiness)" (p. 80). Corresponding to the second criterion are the other three tests. One is "whether one could really be doing it without actually saying anything" (p. 79). Two further *verbal* tests which are aimed at the same notion are: "whether it makes sense to say 'Does he really?'" and "whether

we could insert before the supposed performative verb some such adverb as 'deliberately' or such an expression as 'I am willing to'" (p. 80).

Austin does not say or imply that these tests are all inescapably vague and indecisive. The utterance of some formula might on one occasion pass the tests for a pure performative and on another fail, or come out as a pure descriptive. But this does not entail that its status on either occasion is in doubt. However Austin would, I think, hold that the fourth test cannot be satisfactory, since he has already argued (in my view unsuccessfully) that the notions of truth and falsity are not sharply distinct from the notions of happiness and unhappiness.

The other tests are still vaguer than this. And this goes with the vagueness of the idea of "doing something," an *action*. Austin gives "being sorry" as a possible instance of this idea, which shows how widely he took it, and shows how unclear is the whole idea of "action" he works with in these Lectures. This means that these tests cannot be applied with clear and definite results. Anyway, one might suppose that the idea of "doing something" is *essentially* vague, so that *in a certain sense* it admits of the phenomenon of a "shift." Being sorry is one case, and sneezing is another. (See the essay "A Plea for Excuses.") So the notion of a performative, insofar as it is defined by the idea of "action," would itself be essentially vague. And in *this* sense Austin's claim that there is sometimes a "shift" from performative to descriptive and a "wavering" between them would be justified. But in *this* sense only.

But the tests do not even seem to work clearly in every case. If someone says "I promise," why cannot we ask whether he really promised? And why cannot we say such a thing as "I am willing to promise"? So even promising does not seem clearly to be a pure performative!

With *verdictives* we get a novel consideration. For when a judge says "I hold that," then his saying this is his holding that But when a "less official" person says this, his utterance may be merely descriptive (see p. 88). If the utterance conforms to the initial conditions for the happiness of a performative, it counts as a performative; but the same form of words can be used, *in other conditions*, as descriptive. However, this does not blur the performative-constative distinction. We can clearly distinguish being a judge from not being one, and a judge's attending a session of court from his not doing so, etc. What it shows, though, is that we

cannot exclusively classify *verbs* as *either* names for conventional performances *or* names for (non-conventional) activities or states of mind, i.e., those mentioned in the constative. Nor can we even say that the first-person singular present indicative active is a mark of the performance, other forms marks of the constative. But we may still distinguish *particular* utterances as *either* performative *or* constative.

This kind of case shows how the same form of words may on one occasion pass the tests for the "pure" performative and on another fail them. For, as said by a judge in the right circumstances, we cannot ask "Is that true or false?", etc. But, as said by someone else or in other circumstances, this question and the others are in order. Further, it shows that in trying to decide whether these questions can be asked or not, we have to take into account the particular circumstances of the utterance. So it is not the words "I hold that *p*" that are in themselves true or false—or not. They can be true or false as spoken in some circumstances, but not in others.

With *expositives* we get a final consideration. "Here the main body of the utterance has generally or often the straightforward form of a 'statement,' but there is an explicit performative verb at its head which shows how the 'statement' is to be fitted into the context of conversation . . ." (p. 85). Austin refers to this later as a case where "the whole utterance seems essentially meant to be true or false despite its performative characteristics" (p. 89). The most obvious examples of this phenomenon are the pure explicit performatives like "state" and "maintain," where "surely the whole thing is true or false even though the uttering of it is the performing of the action of stating or maintaining" (p. 90). This seems to be the consideration that is presented as finally justifying a complete rethinking of the performative-constative distinction. For these kinds of utterance seem to fail *both* original criteria for the performative. They are "true or false," and they do appear to "describe, report or constate" something: they do seem to be "statements." But for all that they are, or are parts of the *doing* of some *action*: they look like explicit performatives.

But even if this characterization is correct, the existence of such utterances does not cast any doubt on the possibility of distinguishing a class of utterances which conform to *both* original criteria for performatives. The sharp distinction between "I promise" and "The cat is on the mat" is not

blurred by the existence of a class of utterances which fail to conform to the criteria to which these conform. Rather, the existence of such utterances casts doubt on the necessity of a *connection between* the two original criteria. It was perhaps implied there that performatives are neither true nor false *because* the uttering of the words is the doing of an action. And if that was implied, Austin is right in rejecting it.

This argument casts doubt *not* on the existence of *some* distinction between performatives and constatives, but on some ways of making the distinction. It casts doubt on trying to distinguish them by saying that performatives are, but statements are not, the doings of actions. This, no doubt, is why Austin concludes that it is time to grapple with the fact that making a statement, too, is "doing something."

But the argument also casts doubt on an assumption about truth and falsity which Austin perhaps wants us to believe underlies his original performative-constative distinction: namely, the assumption that truth and falsity cannot belong to "actions." For, if it is correct, we must conclude that the claim that statements are both "true or false" and (involved in) actions is paradoxical, if not actually self-contradictory. This claim is important; but it is not really paradoxical. It only looks paradoxical if we confuse the *act* of making the statement with the *statement* that is so made, i.e., *what* is said. It is *what* is said, the statement, that can be true or false; the act of course cannot. And it is *making* the statement that is an action; the statement itself is not. Austin's conclusion looks paradoxical only if we think he has shown that a certain class of *actions* can sometimes be "true or false." This is nonsense, and he has not shown it. I say this because I am not quite certain what he thought his arguments proved and (unhappily) not quite certain that he made this obvious distinction.

There is a seeming clash of doctrines at one important point in this argument. First Austin claims "surely the whole thing is true or false" (p. 90). But in the next paragraph he refers to "the bit in the that-clause which is required to be true or false." Surely it is the latter which is correct. When someone says "I state that *p*," what is true or false is not the proposition that the speaker states that *p*, but the proposition that *p*. And the sense in which "the whole thing" is here true or false is not a sense which conflicts with this. For it is a sense in which the whole thing will be true or false

even when a man says "I promise" (for example): the sense in which the proposition *that the speaker promises, etc.*, is true or false. Austin's first claim that it is the whole thing that is true or false perhaps makes it easier to confuse the whole thing which is *what* is said with the *act* constituted by saying the whole thing.

Austin's point can be made much more simply, as he in effect goes on to make it, by considering ordinary statements like "The cat is on the mat." Here the whole thing is just what is meant to be, and is, true or false; and it has performative characteristics at least in the sense that *saying* this is *doing* something, and perhaps also in the sense that an explicit performative verb (e.g., "state") could be used to make clear *what precisely* the speaker was doing. But this is not at all odd or paradoxical. It is just to point out that it is possible to *say* (*state, urge, etc.*) *something* that is true or false.

Conclusion

I have tried to show that none of the arguments Austin uses in the first part of his Lectures suffices

to prove the conclusions he seems to want us to draw. He does not show convincingly that the distinction between true-false and happy-unhappy is blurred or non-existent. He does not show that his original two criteria fail to mark out two distinct kinds of utterance. He does not show that there are any overridingly important logical similarities between these kinds of utterance. He does not show (what perhaps he wanted us to believe) that the distinction between *saying* what you say and *what* you say is not clear and sharp. Generally, his arguments do not show that the "special theory" of performative utterances breaks down or that it *must* be reconstituted within a general theory to be viable. And, at least in this part of the Lectures, he fails to show that the traditional view of "statements" is untenable. So we need reject *neither* this traditional notion of the statement, *nor* Austin's own valuable notion of the performative and its "happiness or unhappiness." And even if the general theory of speech acts seems worth attention, we should not feel that his arguments force us into accepting it at the expense of the special theory of performatives.

McGill University

Received September 12, 1968

VI. PERSONAL IDENTITY

CHARLES B. DANIELS

THE classical problem of personal identity is, roughly, that of whether bodily continuity is a logically necessary condition of personal survival. The aim of this paper is to provide a summary and critical evaluation of the more compelling arguments that have been put forward *pro* and *con*¹ and to add a few more.

Consideration of a case, whose source I don't know but which has circulated widely among philosophers by word of mouth, serves to illustrate the in's and out's of these matters admirably:

A machine is built that will, when a person enters it, record the type and position of each molecule in his body and then disintegrate him. The process takes a few seconds and ends with a pile of debris on the floor of the recording chamber. The tape containing the recorded information can then be fed back into the machine; and after raw materials are added, the machine will fabricate a person who not only looks and talks exactly like the one who first entered the machine, but also believes that he is that person. No one who emerges from the machine complains of any suffering he has undergone.

People become accustomed to "travel" *via* these machines. They walk in, are taped and disintegrated, the taped information is beamed to their destinations where other machines await, and there they are reconstructed. "Operations" are performed that were not possible before: a technician makes a few emendations in the tape before feeding it back into the machine. Broken bones are mended, and happier "memories" are furnished for those who simply cannot cope with their real ones.

Here an objection might be raised. Since it is possible for a machine to record its information in duplicate, duplicate tapes can be fed into two machines and there will emerge two people who both claim to be the same person. Now it is in one way possible for the same person to be in two places simultaneously: by standing astride the doorway be both inside and outside his house. But one person cannot, it seems, simultaneously be in two places

that are not connected by some path of places all of which have him situated in them. This is what is normally meant by saying that a person cannot be in two places at the same time. Now if the two individuals that emerge from the machines are the one person they claim to be, then one person *can* be in two places at once. But one person can't. So they both can't be the person they claim to be. But to say that one of them is and the other isn't the person who first entered the machine is arbitrary in a case like this, at least more arbitrary than saying that neither is. So that is the best thing to say: neither of them is. And since it is always possible to make two people from a tape, these considerations hold even when in fact only one is made.

But what this objection fails to note is that the two people who emerge do not claim that they *are now* the same person. They can see very well they aren't. What each claims, on reflection, is that he *was* the fellow that entered the machine, say, an hour before. So both *can agree* (i) that they are now different people and (ii) that they were once the same person.

Other uses of the machine can now be envisaged. Two women fall desperately in love with me. I walk into the machine and duplicate tapes are made. The tapes are fed into two machines, and both of my admirers and "I" live happily ever after. But all is not sweetness and light. Criminals can now duplicate themselves a hundredfold after committing the most heinous of crimes. Are the courts to condemn them all—especially those that reform during the lengthy pre-trial period?

There is another objection to consider, however. A horrible rumor has begun to circulate about the entrepreneurs who control these machines. The machines have only some of the capacities advertised. They can record the information, and they can, given raw materials, fabricate people, but they cannot perform the disintegration operation. What really happens, once the tape is made, is that a trap door opens in the bottom of the recording chamber,

¹ Many of the points that will be discussed here can be found stated or at least hinted at in two articles by B. A. O. Williams: "Personal Identity and Individuation," *Proceedings of the Aristotelian Society*, vol. 57 (1956-57), pp. 229-252; "Bodily Continuity and Personal Identity: a Reply," *Analysis*, vol. 21 (1960), pp. 43-48.

and the unsuspecting soul inside falls into the basement, where he is seized by a band of sadists, bound, tortured until nearly dead, and at last thrown half alive into a vat of acid in which he dies and is quickly decomposed leaving no trace—and all the while on the floor above the accomplices of these fiends are making a duplicate. One can hear them: "How's it going down there? Boris is almost ready to pull the double" will come the shout from above, and from below "Oh, we've all had our fun. He's in the bath."

So the duplicate is *merely* a duplicate, like a photocopy, and not the original. And why do we say that the duplicate and original are different people? Because a person ceases to exist when his body does—by dissolving in acid, or by being disintegrated by a machine—even though, like any item of mass production, there may be many others just like him. Since lack of bodily continuity between the person entering the machine and the person emerging seems to entail that they are different people, we might conclude that a logically necessary condition of personal identity over time is bodily continuity.

That they are different people is in a case like this the correct answer, I think, but another consideration seems to suggest that our judgment to this effect is not based upon seeing an entailment, i.e., bodily continuity is not a *logically* necessary condition of personal identity over time. Cases of reincarnation, if such things are logically possible, are precisely cases in which there is personal identity in the absence of bodily continuity.

Suppose my oldest friend awakes one morning thoroughly convinced that he is Henry Morgan, the pirate. I know little of Morgan, but from what I do know I judge that my friend has taken on the personality and character traits of Morgan. But this may simply be due to an unconscious desire on his part to preserve a deluded conviction that he is Morgan; to keep in character he must behave like him. In any event, people can have the same character and personality traits and still be different people, so having these traits certainly isn't sufficient to make my friend be Morgan.

But he also seems to know vastly more about the intimate details of Morgan's sordid life than the John Dokes I have known could possibly be in a position to know. As boys we played pirates, but I know for a fact that he took no more than the passing interest I did in the lives of pirate notables. He insisted, moreover, in making a trip to the Caribbean "to convince you sceptics"; and near

Ocho Rios, after having oriented himself to the altered landscape, he pointed to a spot where we dug up a treasure and the bones of the men he claimed he had murdered to preserve the secret of its location.

If we decide that Dokes is a reincarnation of Morgan (or that Morgan's soul has informed Dokes's body), we do so having judged that the claims Dokes makes that he *remembers* being Morgan, burying the treasure, etc., *are true*; it is a real case of remembering, and this judgment forms the basis of our decision that Dokes is Morgan. However, our basis for judging that Dokes *remembers* being Morgan, burying the treasure, etc., is that we have good reason to believe that *Morgan* buried the treasure, killed his crew, and whatever else Dokes claims *he* did, and that Dokes has some, to say the least, very extraordinary *insight* into Morgan's life and deeds.

But now another objection must be raised. We might account for Dokes's strange insight in other ways than by calling it remembering. Dokes might simply have clairvoyant knowledge of what Morgan did, felt, thought, etc., and be mistaken when believing it to be true he says he *remembers* being Morgan, burying the treasure, etc. One can think one is remembering events which one is, say, merely imagining, and one can think one is merely fantasizing events that one is really remembering. So, too, it seems possible to have clairvoyant knowledge of something one *only thinks* one remembers.

Suppose I claim to have a wonderful clairvoyant power to know things. When asked to give examples of things I know this way, I say "I know that I had eggs for breakfast this morning." This isn't clairvoyance. I was present when this happened and am remembering it. We know this by knowing that the body I have was in a position for me to be seated at the breakfast table and to know that I was, etc.

It seems, then, that bodily continuity plays an essential role in distinguishing cases of clairvoyance from cases of remembering. This is undoubtedly true. It might be argued further, though, that when bodily continuity is absent as it is in cases of reincarnation or "change of bodies," the distinction between the two collapses. Remembering is one thing when it is distinct from clairvoyance; it is quite something else when it is not.

If this line of reasoning is accepted it is suggestive of further argument. First, that if "remembering" doing something in a former life entails that one

did it, the concept of *person* we are dealing with differs significantly from our normal one. For example, no present court would adjudge Dokes guilty of the murders of the crew members even if it accepted that he "remembered" committing them and, hence, was the "person" who did—nor, I imagine, would it judge that he was not married to Dokes's wife if it accepted that prior to his awakening as that "person" he was not that "person"—although this change might serve as grounds for divorce. And although we might find much objectionable in his present character, even viewing him with suspicion as a potential criminal because of these traits, we would not, I think, regard him with the same moral disapproval as we would if he had committed murders in his present life and remained unrepentant.

Secondly, there seems to be no reason to hold that the implications of "remembering" (which is not distinct from clairvoyance) will resemble those of our normal remembering (which is). In particular, the fact that remembering that one has done so-and-so entails that one has done it when the presence of bodily continuity preserves the distinction provides no clear reason for holding that the entailment obtains when the absence of bodily continuity collapses it. So the fact that Dokes "remembers" killing the crew is no longer a reason for saying that he did it. It may still be a reason for saying that someone did, for remembering-clairvoyance still has some element of knowledge in it.

But to all this a believer in the possibility of reincarnation will just reply that *there is* a distinction between remembering and clairvoyance despite the fact that in such cases bodily continuity is absent: if Dokes remembers killing the crew and if Morgan singlehandedly killed the crew, then Dokes is Morgan; if Dokes merely has clairvoyant knowledge that Morgan killed the crew, he isn't Morgan. The fact that we do not seem to be able to figure out a way of coming to know which is in fact the case does not imply that there is no distinction between the two kinds of cases; nor does the possibility of two or more people making whole batteries of sincere claims which if true would necessitate their being Morgan imply that it is impossible for one and only one of them to be Morgan.

Yet the question does remain of how we can tell the rememberer from the clairvoyant. If in normal cases bodily continuity is treated merely as an inductive ground for saying yea or nay, there must be an independent way of telling which is which if

we are to establish empirically that it is a good inductive ground, or so it seems. It might be that we just see that *X* is remembering and that *X* isn't. But if this is our answer, why shouldn't we simply forget about remembering that one is so-and-so, a *sufficient* condition of identity, and just say that we see that *X* is so-and-so and the *X* is not? If the answer takes either of these forms, however, we can still perhaps be jollied out of it.

Suppose I am an anthropologist who has been living with a tribe long enough to have sufficient command of their language to carry on a fair conversation. One day the chief and I are sitting and watching archery practice. One of the men is particularly inept, invariably missing the target completely. I remark to the chief "He's certainly not much of an archer," and to my surprise the chief answers "Oh, he is in the next village." I ask "Do you mean that he becomes a good archer when he goes to the next village?" and the chief replies "No, I mean that he is *now* a good archer there." I ask "Do you mean that he is in the next village now, besides being here too?" and the chief answers "Yes, and what's more he is sitting over there too, but that's not he over there," pointing in the first instance to a woman nursing a child and in the second to a teenager. It seems clear, if the old boy isn't pulling my leg, that these people have a different concept of a person from ours.

I then ask the chief how he knows that the man in question is in the next village and he replies "I see that he is. I'm there too." I ask him how one tells which of these figures are the same person and which aren't, and he says "Well, it's just one of those things, like sameness of color, that one sees, or one doesn't. A few of us are 'person-blind' and can't do it. There have been cases where a foreigner who has settled here and learned our ways has finally gotten the knack of it. You may be suspicious, but if you remain with us, you will find that you have good reasons to believe that we can see these things, just as blind people have good reasons to believe that those who claim to have the power of seeing colors actually do see them. Blind people know that there is overwhelming agreement about what colors things are. They know that this agreement cannot be traced to authority; you don't need traffic policemen standing on corners shouting 'Now the light is red. You must stop.' They know, moreover, what organ would have to be fixed for them to have this power. It is the same with you. You will find that we agree about what things are and aren't the same person. Our agreement doesn't

come from authority. And your eyes may well need fixing."

Now the person who believes in reincarnation might not be convinced that his concept of remembering or of person differs from the normal simply because he realizes that this tribe has a different concept of person from ours. But the chief's argument does have a sufficiently strong cajolery force to make it hard for him to say that he simply sees the difference between remembering and clairvoyance or between the same and different people and that the rest of us are blind in these respects. And there is the further difficulty that our world is full of instances of the same color, just as the tribesmen's is full of instances of the same "person." Cases of the Dokes type of insight just do not occur, so questions of whether one would be driven to take the "simply seeing" line if they did become difficult to answer without further information.

This is an important point and it can be illustrated by the difference between the machine case and the case of reincarnation or "change of bodies." When we read the details of the machine case, we find they fit easily into a whole set of background beliefs we have about how things actually work in the universe. The people emerging from the machines have the personalities and beliefs they do *because* personality and belief is determined by certain molecular configurations in one's body—at least that is *our* belief. And when we realize that the machine-constructed people are not the ones who first entered, we are *still* able to give this explanation of why they behave as they do. We do not *have* to fall back upon such outlandish things as clairvoyance or queer rememberings to explain it (not that a thing cannot be explained by both the concept of molecules at work and by the concept of remembering). But if we didn't have the background beliefs we do, the machine case might be on a par with the reincarnation case, except for the inclusion of a lot of unnecessary trappings.

Why do the people act, talk, and believe as they do? Because they have certain kinds of molecules in certain parts of their bodies. Why does Dokes behave as he does? Same reason. Why are these molecules there and structured in just that way in the bodies of the people who emerge from the machines? Because someone by means of the machines put them there and arranged them that way. Why are the molecules there and arranged in just that way in Dokes's body? We just don't know,

not that we can't. Perhaps Morgan's spirit informed his body, causing a molecular rearrangement. Perhaps a combination of clairvoyant power and mental instability made him that way. Perhaps molecular configuration has nothing to do with one's mental states. Imagine a case of the machine or John Dokes type happening in an *Oz*-like world of tin men and straw men in which molecules had nothing to do with what its inhabitants did or thought. What would be the right thing to say here?

I think the difficulties in deciding what would be the right thing to say arise from two factors: (a) a suppressed demand for physical possibility and (b) lack of relevant information. Another case may help. Sydney Shoemaker offers the following counterexample to the thesis that bodily continuity is a necessary condition of personal identity

... suppose that ... a surgeon can completely remove a person's brain from his head, examine or operate on it, and then put it back in his skull ... without causing death or permanent injury; we are to imagine that ... "brain extraction" has come to be widely practiced in the treatment of ... disorders of the brain. One day ... a surgeon discovers that an assistant has made a horrible mistake. Two men, a Mr. Brown and a Mr. Robinson, had been operated on ... and brain extractions had been performed on both ... At the end of the operations ... the assistant inadvertently put Brown's brain in Robinson's head, and Robinson's brain in Brown's head. One ... immediately dies, but the other, the one with Robinson's body and Brown's brain, eventually regains consciousness. Let us call the latter "Brownson." Upon regaining consciousness Brownson exhibits great shock and surprise at the appearance of his body ... Pointing to himself he says "This isn't my body; the one over there is! When asked his name he automatically replies "Brown,""

and, in sum, thereafter acts like, talks like, and believes himself to be, Brown. Shoemaker himself has doubts about what is the right thing to say. I shall limit myself to commenting on the case as it is presented above.

It seems to me, right off, that Brownson is Brown. I am inclined to say this because he has Brown's brain and behaves as the story says. Brain continuity is not, however, a logically necessary condition of personal identity.

Shoemaker goes on to argue that if:

upon regaining consciousness Brownson were to act and talk just as Robinson had always done in the past, surely no one would say that this man, who looks, acts, and talks just like Robinson, and has what has always

² Sydney Shoemaker, *Self-Knowledge and Self-Identity* (Ithaca, Cornell University Press, 1963), pp. 23-24.

been Robinson's body, must really be Brown rather than Robinson because he has Brown's brain. Here we would conclude simply that there is not the close causal relationship we had supposed there to be between the state of man's brain and his psychological features, i.e., his personality and his ability to report events in his past history. If we did not think there to be such a causal relationship we should not think that having the same brain has anything more to do with being the same person than, say, having the same liver.³

If Brownson were to awake acting and talking like Robinson, I myself do not think we would immediately conclude that brains do not have the intimate relationship we think they have to psychological features. There is simply too much other evidence that they do, e.g., the effects of frontal lobotomies and brain tumors, etc. Brownson's case would become an anomaly, and many many other alternative explanations would have to be explored before we would resign ourselves to saying that as regards an explanation of human consciousness brains play about the same role as livers and that these other "effects" (of frontal lobotomies, etc.) are sheer coincidence. Suppose, for instance, that Brown has had a history of clairvoyant interludes and also has shown signs of mental instability. Immediately before his operation he had heard that someone named Robinson was undergoing a similar operation. Perhaps he is now fulfilling a wish to be someone else and picked on Robinson. This is farfetched, but it seems no more farfetched than the supposition that Brownson wakes up believing that he is Robinson and acting like him, given that he has Brown's brain. Equally farfetched would be the case in which Brown awakened thinking and acting like Robinson after receiving a transfusion of his blood.

What all of this shows is, I think, that we have a natural tendency, when presented with cases like these, to fill in the missing details in such a way as to make the imagined world, the one we have in mind, as much like the actual world with respect to unspecified detail and physical possibility as we can, given the limitations of the statement of the case and our beliefs about the actual world and physical possibility. If, for instance, we harbor a belief that there is an identity *in fact* between mind and brain, we will take it for granted that there is this identity in a case like Shoemaker's and say that Brownson is Brown. Robinson's brain now lies encased in the inessential part of Brown's body—

inessential in the way fingernails and livers are inessential, while the essential part of Brown's body, which is in fact his brain, lies encased in the inessential parts of Robinson's body. (Suppose that brains could be fabricated by a machine so they exactly resembled our own brains and suppose also that they could be transplanted in our skulls in an operation like the now successful kidney and heart transplants. But remember the possibility that sadistic fiends might be in control!) Or if we use the word "body" in Brown's, perhaps more normal, way, we shall say that despite the fact that one's bodily continuity is not essential to one's survival, *spatiotemporal continuity of something*, whatever one cares to call it, may still be; and this something in the imagined world *and* our own is in fact the brain. And because we believe this we say that Brownson is Brown.

But suppose that further details are given: suppose Brownson's opposite number, call him Rown, doesn't die, but awakens *also* thinking, talking, and acting like Brown—we don't know what to say because we don't know who is who. And we don't know who is who, because now, it seems, we don't know enough relevant details. But the tendency to make the presented world as much like what we believe the real world to be within the framework of detail actually given still operates, and this determines what kinds of further details we feel ought to count as relevant. Given a belief in the identity of mind and brain, the case becomes mysterious, because we want to reconcile this belief with the story told. Then when further details are given that make it clear that brain-mind identity does not hold in the hypothetical world, other beliefs about the real world still retain their hold: we believe that a person could not be such a lucky guesser that he would with no knowledge about the life of another guess all the details of that other's life with the facility and accuracy that the other could remember them, that he would not have this detailed clairvoyant knowledge, that if he were guessing or clairvoyant, he could not be so mistaken about his own identity, that he could not be so mistaken about his mental states, that he could not react in just the same way as the other to people and situations, that in all probability there are no cases anyway of clairvoyance, reincarnation, or one soul, mind, or what have you, informing different bodies. Any of these beliefs might be false; none of them are, in my opinion, logically necessary truths.

Here I am handicapped by the fact that I have

³ *Ibid.*, p. 24.

no analysis of knowledge or belief to offer, so what I say in the following has no other basis than my own intuitions. I ask the reader to test mine against his. Let us operate under the assumption, though, that the worlds presented in these cases differ from the actual world, represented partially by the truth of the beliefs stated above, only as much as the details stated in the descriptions of the cases and taken in a generous spirit require.

In the original brain-transfer case, we should, I think, barring sheer stupidity, *know* that Brownson was Brown.

In the case of Dokes-Morgan, I think we might have good reason to suspend judgment, to wait and see what turned up. In fact, if no other such cases occurred, generations in the fairly remote future would probably have good grounds to doubt that the case of John Dokes had ever really occurred, unless investigations into the mind-brain relationship take a very surprising turn.

But if such cases were to begin to occur frequently, provided, of course, only one good candidate appeared each time, we would have good reason to believe—and we might perhaps even come to know—that Dokes was Morgan. The mind-brain identity thesis would have to be rejected, but minds could still be thought of as having an exceptionally intimate causal relationship to brains during their incarnate periods. Also, to be sure, we would start believing that cases did occur of minds informing different bodies at different times. Our reasons for believing Dokes to be Morgan would, I imagine, be that we knew that in all normal cases the likelihood that memory was the source that provided a vast store of true information about the past doings and thoughts of someone is far greater than the likelihood that it came *via* clairvoyance or sheer uninformed guessing. Also, we regard it as wildly unlikely that a man could be mistaken on so grand a scale about whether he was remembering, being clairvoyant, or just guessing blindly. And we could still say that there were no cases of clairvoyance. Of course, the problem of determining the mind-brain relationship would become far more difficult for the scientists, perhaps requiring a new physical theory.

Similarly in one isolated case of competing candidates, whether it is a case where no spatiotemporal continuity is apparently present or one like the case where both Brownson and Rown act like and believe themselves to be Brown, judgment is best suspended.

We do want to hold it a logically necessary truth,

I think, that in the absence of physical continuity *at most one* of the competing candidates is who he claims to be, i.e., two such candidates cannot have been one person in the past. The qualification as to spatiotemporal continuity is required by another kind of case: if children were born fully able to reason and talk, and each firmly believed it had done and thought the things its mother had done and thought, we might want to say that the mother and child had once been the same person. Leaving this aside, though, we realize that it is logically possible for duplicates to exist, machine made, for instance. But once this is pointed out, we also realize that the duplicates aren't and weren't the same person. Saying that two people *were* the same "person" in absence of spatiotemporal continuity is too much like saying that they are the same "person," i.e., that a "person" can be in two places at once. Far too many of our customs and values depend upon the fact that this cannot happen: monogamous marriage, the institution of promising, testifying, and calling into question testimony, etc. Vastly different cultures might not find our concept of person useful and might employ another.

At any rate, where it becomes the regular thing for equally good competing candidates to appear, our whole stock of basic beliefs are, considering the case's intent, called into question. Whether or not minds are informing different bodies, there are cases rampant of clairvoyance, incredibly lucky guessing, being mistaken on the grand scale about one's mental states, or something else. As to the "something else," perhaps there is a powerful Martian causing these things to happen, making some minds inform now one brain, now another, causing others to believe that they are yet other minds and filling them with information—perhaps by causing changes in brains. And the Martian might even be clever enough to be able to explain to us how he does it.

But this sort of explanation of what is going on is not meant to be available. On the one hand, our background beliefs and sophistication about the actual world don't help; and on the other, we are not meant to be allowed to develop a new set to replace them. So by the very nature of the case, if we are to know which are "change of body" cases and which are not, we seem to be reduced simply to seeing that so-and-so is actually remembering and that so-and-so is being clairvoyant or guessing blindly. It is good to note, however, that in some possible world people could have full-blown and sophisticated sets of background beliefs about these

matters, about why they happen and what mechanisms are at work, rather than having to view them from eyes, very much unlike ours, that never look for explanation but merely identify. But from such eyes, spatiotemporal continuity is not an aid for making the distinction; and if criteria are demanded (like those we could perhaps name for a thing's being a unicorn), we fail to come up with them.

It is a mistake, though, to become mesmerized at this point by the case of colors and to turn it into a paradigm for all cases in which we simply see that so-and-so is the case without being able to name interesting criteria. I believe that seeing that a thing is changing also has this feature. Wittgenstein suggests other cases. What are the criteria for a thing's having a gentle, benign, or stern look? Talented artists can capture these features on canvas or on stage, but we and they are at a loss to name criteria for them (unless, to be sure, gentleness is treated as its own criterion).

What makes a thing a gentle look seems far more complicated than what makes it, say, red. Parts of a wholly red surface can be removed and the remainder will still be red. Parts of a face that has a gentle look cannot so easily be removed leaving the gentle look intact. If a man has trouble making these distinctions in practice, his problem seems much less one of the organs of sight than one of sensibility, the cure less amenable to drugs or operations than to training and even perhaps change of character. The distinction is not sharp, of course; wine tasters develop extraordinary sensitivity, as do actors.

My suggestion is that the distinction between remembering, being clairvoyant, and lucky guessing might be viewed the same way, and what makes a thing a case of one rather than another is perhaps even more complicated than its having a gentle, benign, or stern look. We are handicapped, too, by the fact that there simply are no cases of clair-

voyance and such wildly lucky guessing. It is as if there were only one flavor of tea in the world, and we were asked to imagine a world in which there were others.

In summary, I think that the believer in the possibility of reincarnation or bodily transfer is right on logical grounds in insisting that there is a distinction between remembering, clairvoyance, and uninformed lucky guessing even if bodily continuity is absent. I fail to see why there would cease to be a distinction even if it were logically impossible for us to know which was a case of one, which of another—much less if it were merely impossible in practice. But I have also suggested that it *would* be possible for us to know which was which and even to explain the hows and whys of the phenomena. I have also suggested that we do distinguish many things of high complexity, unlike color shades, without being able to say, except trivially, what the distinguishing marks are, and that the present distinctions might be of this kind.

But in judging whether we are still talking about the same thing by the word "person" in describing a world where competing candidates occur frequently, other details would be relevant: Does change of sex ever occur when one mind informs different bodies? Does one have to obtain a divorce from someone who has changed bodies? Is one guilty of crimes committed when one had another body? Is one's eye-witness testimony accepted in court after a change of bodies? Can one's named heirs inherit when one dies? Do minds survive in space invisibly when one dies? Do the scientists of the world attempt to test for the presence of disembodied minds in space? I would imagine, here, that the answers to such questions could be filled in in such a way as to make it fairly plausible that we are still using the word "person" in the regular way and that change of bodies does occur in the absence of any form of spatiotemporal continuity.

VII. LANGUAGE AND COMMON SENSE

JAMES E. BROYLES*

IF one is asked to characterize George Edward Moore as a philosopher two things spring to mind. He is associated with Wittgenstein as one of the two powerful forces that have shaped the so-called ordinary language movement, and he is thought of as the principal 20th century advocate of common sense.

Now there is a certain tension between these characterizations. The former would seem to demand a fair measure of continuity between Moore's work and that of those who followed him. Yet the latter points toward the uniqueness of Moore's method, for it must be admitted that the appeal to common sense is at least *prima facie* distinct from the appeal to ordinary language, and that no philosopher of the generation influenced by Moore has made extensive use of an explicit appeal to common sense. Does Moore represent a genuine anomaly in the analytic tradition? This is one of the questions we might answer in reconsidering the relation between those two very basic appeals, to ordinary language and to common sense.

Now there have been philosophers radical enough to deny that there is any legitimate appeal to common sense over and above the more familiar appeal to language. Malcolm,¹ for example, insists that the real force behind what Moore does, whether Moore realizes it or not, depends upon the same principles of ordinary language as those that subsequent philosophers have relied upon. This may seem to be stressing continuity with a vengeance, since it would virtually deny that there is anything of value in those aspects of Moore's method which are unique to Moore.

Persistent critics of Malcolm on the other hand (White,² Chappell,³ and others), have shown the plausibility of an alternative account of Moore's appeal to common sense, one which depends upon truth rather than upon language. They maintain that there are two independent appeals, and they

hold that this is shown by the fact that there are two different ways in which one might attempt to justify what Moore does. Perhaps the danger here is in making Moore's technique appear rather more autonomous than it really is.

In this paper we shall not be primarily concerned to defend or justify either the appeal to common sense or the appeal to ordinary language. We intend to examine certain aspects of each and attempt to show the intimate way in which they are inter-related. Our goal is to show how this bears on the understanding and appreciation of Moore's work.

I. PRELIMINARIES

Let us begin by drawing attention to the distinction between the constructive reflection upon language and the destructive appeal to ordinary language. When a philosopher reflects upon language, he tries to promote an awareness of the specific features of the context within which an expression is correctly employed. He reminds us of the characteristic circumstances under which the expression is normally used. Now he may do this for a variety of purposes. He may be primarily interested in the language itself, or he may be interested in a contrast between the familiar use and some technical use, or in the eccentric speech of some individual or group of individuals, and so on.

Such reflection upon language has never been the monopoly of any school of philosophy and it would not in itself appear to be open to serious objection. There is always the question as to whether it will be fruitful or not, but there is no serious challenge of its legitimacy. Even the question as to whether there is any familiar use of an expression would appear to be one that is to be answered by such reflection rather than constituting an objection to it.

It is only if a philosopher goes on to argue that

* The author gratefully acknowledges the award of a Summer Stipend (67) by the National Foundation on the Arts and the Humanities. The present paper is largely a product of the study supported by that grant. A draft of this paper was presented as the Presidential Address at the Northwest Conference on Philosophy, April 27, 1968.

¹ "George Edward Moore," *Knowledge and Certainty* (Englewood Cliffs, 1963), pp. 163-183.

² G. E. Moore: *A Critical Exposition* (Oxford, 1958).

³ "Malcolm on Moore," *Mind*, vol. 70 (1961), pp. 417-425.

other philosophers are wrong *because* they have misused language that he can be described as appealing to, or arguing from, ordinary language. It is this destructive use of ordinary language that has served to identify a movement. It is with this use that we will be concerned. (The signs of the times would seem to justify us in referring to the ordinary language movement in the past tense.)

Turning to Moore, the first thing that we might wish to notice is that he rarely makes destructive use of ordinary language. He seldom criticizes a philosopher on the grounds that he has misused language, seldom rejects a thesis on the explicit basis that it is not in accord with the standard use. Here we have the apparent divergence between his work and that of the early practitioners of ordinary language.

This divergence is sometimes obscured, however, by the fact that Moore does make dramatic and powerful use of "what we should all naturally mean," "what we should all understand by," and so on. He engages in constructive reflection upon the language. He does so in his efforts to clarify the meaning of various philosophical claims, in his attempts to show what a philosopher intends when he says what he does say. It is only if one treats any reflection upon language as an appeal to ordinary language that one can make any considerable case for there being two wide-spread appeals in the work of Moore, as some philosophers, White for example, have tried to do.

In discussing Moore's technique we are led to consider three distinct features. We shall refer to them as (1) Moore's analysis, (2) Moore's end-game, and (3) Moore's principle.

The first task for Moore is always to find out exactly what is meant by, say, the doctrine of intrinsic relations, or the assertion that time is unreal, etc. We are calling this Moore's analysis, though of course for Moore the term "analysis" referred to no such preliminary activity. Rather it referred to that elusive end-product of successful philosophizing. This early notion of an analysis was soon rendered suspect by its conspicuous absence in the actual practice of philosophy.

At any rate, it is in Moore's painstaking dissection of philosophical issues that he remains true to his own well-known observation that in such studies "the difficulties and disagreements . . . are mainly due to . . . the attempt to answer questions, without first discovering precisely what question it is which you desire to answer."⁴

⁴ *Principia Ethica* (Cambridge, 1903), Preface p. vii.

It is here that Moore relies most heavily upon the familiar use of language and it is perhaps here that he seems most modern. Moore's great analytic powers, his standards of clarity, his capacity for fine distinction, his nose for ambiguity; these are the qualities which tend to command contemporary respect.

Even in his analysis, however, Moore also makes use of common sense. But the central point is that while Moore makes use of both what we commonly believe to be true and what we would ordinarily mean in his analysis of the content of philosophical disputes, his refutations of other philosophers, are almost always dependent upon common sense. It is for this reason that it is convenient to distinguish between his preliminary analysis and his end-game. Now just as it is with respect to his analysis that Moore seems most modern, so it is chiefly with respect to his end-game that he is to be contrasted with subsequent philosophers. He argues much as they might argue down to the final move. But for Moore the moment of truth comes when the philosophical thesis, now properly understood, can be shown to be in conflict with common sense. This is the ultimate goal toward which he is driving in his analysis.

We should mention one further point, the relevance of which will be clear at a later stage. It concerns Moore's attitude toward ordinary language. As a matter of policy Moore generally chooses to employ language in a straightforward way. He indicates in a number of places that he thinks other philosophers would be well advised to do the same. This is largely based on consideration of clarity, however, not on the belief that a departure from standard usage is in itself philosophically objectionable. But it is characteristic of Moore to treat the ordinary meaning of expressions as unproblematic. He holds that there obviously is an ordinary use (or several), that we all know perfectly well what it is (or they are), and that it would be foolish not to follow it (or them). No great issues are seen to hinge upon the ordinary use of language. Moore's faith in ordinary language might almost be described as precritical. It foreshadows none of the coming "wars" over the status, the character, the legitimacy of the appeal to ordinary language.

II. MISUSE AND LANGUAGE

Before we can compare Moore's common sense end-game with the more usual end-game of

ordinary language it will be necessary to say something about ordinary language.

Unfortunately a good deal of "learned dust" already clouds this rather well-trod path. The giants have passed this way. Nevertheless, taking courage from the narrowness of our present objectives we ask simply: What exactly is a misuse of language? What is this supposed original sin of metaphysics, this occupational hazard of philosophy? I think it would be wrong to assume that the early practitioners had a clear, simple, univocal notion of what constitutes a misuse of language. But their examples do tend to cluster around a specific sort of error which we shall try to exhibit as briefly as possible.

These days we are all accustomed to the recognition that the expression "use of language" is highly ambiguous. But so, also, must be the expression "misuse of language." There are undoubtedly a number of failings, grammatical, linguistic, and speech, all of which might naturally have been called misuses of language. If we are to avoid the recent Babel of indiscriminate epithets, "absurd," "conceptually absurd," "nonsensical," "pointless," "misuse," and so on, some preliminary sorting is unavoidable.

First of all the notion of a "misuse of language" must be contrasted with the notion of a "nonsensical utterance." Perhaps we can vary the usual motif here by focusing on misuse rather than upon language. It is a general conceptual feature of a misuse (of anything) that it presupposes the notion of a normal or standard use. This carries across the misuse of furniture, dishes, tools, instruments, and so on. We can misuse a microscope whereas we cannot normally misuse a twig, not because twigs are inherently non-misusable, but because they normally do not have a standard use, and until they do they cannot be misused.

Thus the most obvious condition for saying that a linguistic expression has been misused is that it must have a normal or standard use from which there can be departures. And even this has been a source of misunderstanding for it may well seem paradoxical that—though any expression may be used, that is, given a use—not every expression can be misused.

We must note too that not every nonstandard use constitutes a misuse, if we use a microscope for a paperweight, for example. For though this is not its normal use, we should not call this a misuse since the notion of a misuse carries with it the

intimation of harm or damage in some respect, perhaps interference with the capacity for normal use. Presumably then, to apply the concept of a misuse to language following the model of "pots and pans" we must bring in a notion such as that of "an obstacle to communication," "a potential source of misunderstanding," or something of this sort. Let us leave this aside for the moment.

Since every expression that is misused has a proper use there must be conceivable (if not actual) circumstances under which it could be correctly employed. It follows that if an expression is misused this can only be because the actual circumstances are not the appropriate circumstances, and this is one of the essential differences between the misuse of an expression and the use of a nonsensical expression. A misuse is what we might call context relative. What is a misuse in these circumstances is not necessarily a misuse in other circumstances, and necessarily not a misuse in some conceivable circumstances. If an expression is nonsensical, on the other hand, there are no circumstances (actual or conceivable) under which it would ordinarily be used. It does not have a regular place in the language. The concept of a nonsensical expression, though important in its own right, is not our present concern.

Another concept frequently confused with the concept of a misuse of language is the concept of pointless speech.⁵ Here again we are dealing with something on a quite different level. The question as to whether or not an utterance is pointless is a question as to whether or not the conditions of the utterance justify the ascription of certain speech acts to the utterer. Is he informing someone, reassuring them, and so on? It is a question of what one is doing in saying what one does say. Clearly the concept of a pointless utterance is also context relative. An utterance which is pointless in particular circumstances may well have point in other circumstances.

But one feature that does differentiate the pointless utterance from the misuse of language is the latter's intimate relation with the concept of truth. The point or pointlessness of an utterance is independent of its truth or falsity. False utterances may well have point, true ones may lack it.

Now as it stands, this is also true of expressions that are misused since the relation between misuse and truth involves the following complexity. It is not the actual truth that is important for the concept of "misuse," but what the utterer takes to be true. Thus it is theoretically possible for a person

⁵ See Norman Malcolm, "Defending Common Sense," *Philosophical Review*, vol. 58 (1949), pp. 616-621.

to be misusing language even though what he says is true. But this can only be the case where he is also making a counterbalancing factual mistake. It is only possible because linguistic and factual mistakes can, so to speak, cancel one another. Suppose a certain individual invariably calls logs "sticks." Should he overestimate the size of a given stick and think it a log, his statement that he sees a stick may well be true, though he is misusing language all the same.

For simplicity then, let us assume that we can establish that the utterer is fully cognizant of the relevant contextual features. He is not making a simple factual mistake of the more usual sort. Then we can charge the utterer with misuse only if the expression is false when understood in this natural and straightforward way.

Of course it is not sufficient that what he says is false because we must allow for the nonstandard, yet non-misuse, cases previously mentioned, i.e., the microscope used as a paperweight example. We must allow that the utterer may be doing something the success of which is not incompatible with his utterance being literally false. The simplest such case is where the person is lying. But there are also the more philosophically interesting possibilities that he is engaged in eccentric speech, using metaphorical or poetic language, employing a variation of language for emphasis or for humor, and so on. Should we fail to allow for these possibilities the charge of misuse can easily degenerate into logic chopping of the rankest sort. The unhappy consequence of taking these charges seriously would only be the impoverishment of our natural language. However, the standard reply is that these nonliteral uses of language are each, as the expression goes, parasitic upon the standard use, the eccentric upon the noneccentric, the metaphorical upon the non-metaphorical, and so on.

At any rate, few philosophers would leap to embrace a defense that consists in the admission that what they have said is not literally true. They would not be much more likely to do this than they would be to welcome a defense that hangs on the admission that they are factually mistaken, or that they are lying. The charge of "misuse" is hardly so heinous that it must be put down at *any* cost.

Let us bring these points to bear on an example that is typical of the so-called paradox propounders. It derives ultimately from a well-known remark of Bertrand Russell, quoted and made use of by Malcolm. Of course we wish to consider it apart from its historical associations.

Imagine a philosopher, let us call him Propounder, deep in thought, struggling with the conundrums of perception. Suddenly he straightens up, a glow of insight lights his face. Staring at the hand before him, he exclaims excitedly "What I really see is only a part of my own *brain*."

First of all it is clear that his utterance is not nonsensical. We can imagine circumstances under which it would be perfectly correct for one to say this. Because of the physiological limitations (the location of the brain relative to the normal span of vision), we would probably resort to an appropriate arrangement of mirrors or television cameras. At any rate this expression has a place in the language though not the place that is being suggested for it. Now Propounder is not making any familiar, commonplace factual mistake. There is no item of information about his hand, the situation, his faculties, which he has overlooked, or which would account for his remark. Propounder would certainly deny that he was speaking metaphorically. He believes that he has discovered something and that what he says is literally true, however paradoxical. Verdict: Propounder is misusing language.

III. ORDINARY LANGUAGE AND COMMON SENSE

We are now in a position to compare and contrast the arguments that are grounded upon ordinary language with the arguments that are grounded upon common sense.

First of all it is well to note that we can picture Propounder as having made either the particular claim, "What I really see *here and now* is only part of my own brain," or the general claim which is supported by the same considerations, "Whenever one looks at *anything* one really sees a part of one's own brain." Moore's reply, of course, must be particular since he is not holding that we never see our own brain, but only that we very often do not and that Propounder does not in the case at hand. If Propounder's claim is presented as a general thesis, and Moore's rejoinder is particular, we preserve a maximum of resemblance to the familiar case of the refutation of a generalization by counterexample. Thus this mode of presentation makes it especially easy to misconstrue what is actually involved. There is still an unbridgeable logical difference in that normally the introduction of a counterexample involves a new consideration, something that was not contemplated in the making of the original claim.

On the other hand, if the claims of both Moore

and Propounder are presented as particular claims we project in sharp relief the image of Moore's rejoinder as an apparently dogmatic counter-assertion. Propounder says he sees his own brain. Moore says that he does not. The head to head character of the disagreement becomes most glaringly evident. We shall follow this latter course.

Moore's end-game goes like this. A philosophical thesis is advanced. Moore makes an appropriate assertion such as "Propounder sees his hand, not his brain" (where this is to be taken in an ordinary way). He holds that both he and Propounder know with certainty that this assertion is true. The truth of the assertion is seen to be incompatible with the truth of the philosophical thesis (again, taken in an ordinary way). Conclusion: Propounder's thesis is false.

The corresponding ordinary language end-game goes as follows. A philosophical thesis is advanced. We reflect upon the circumstances under which it would be correct to assert the philosophical thesis. When would we say "I see my own brain"? The actual circumstances (which are taken as known) are seen to differ from the appropriate circumstances. Conclusion: Propounder is misusing language.

Now presented in this way it becomes evident that these are not independent techniques. You might say they work from different ends. The language is inappropriate for the occasion. Moore asks what language would be appropriate for *this occasion*, while others ask what occasions would be appropriate for *this language*.

Consider the following analogy. Suppose that someone is very much taken with a hat that he has tried on and is simply unable to see that it does not really fit him. We may be able to bring him to recognize that it does not fit in either of two ways. We may find someone else that the hat in question does fit so that he may see the difference or we may find another hat which does fit him so that he may thereby see the difference.

The reason why the philosophical techniques tend to look different is that in each case different things tend to be assumed or suppressed. Moore, as we mentioned earlier, is inclined to take the standard employment of language as unproblematic and beyond dispute. Subsequent philosophers (with the wane of scepticism) have been inclined to grant our capacity to establish basic facts. Both of these are necessary if we are to conclude that a philosopher is misusing language. The apparent divergence then lies in which of these the users of the "rival" techniques are prepared to accept without question.

But surely, one might argue, the *defense* or *justification* of common sense is to be distinguished from the *defense* or *justification* of ordinary language. Well, yes and no. It is true that though the appeals are in essence the same, the justification suggested by each is different. This is a product of the fact that one is normally called upon to justify one's explicit claims rather than one's implicit assumptions, especially where the latter are shared by the disputants. If one says that one knows for certain such things as that this is a human hand and that the brain is in the head, and so on, then one will be called on to justify his right to make such assertions. If one says that particular expressions are ordinarily used in certain specified ways then he will be called on to justify his right to make *these* assertions. So justification points in two different directions. A defense of ordinary language would presumably involve making clear the fundamental role of the standard practice in language. A defense of common sense, on the other hand, would evidently involve showing the fundamental epistemological status of certain observation statements. But though these are genuinely different sorts of things, *both* would be required if the assumptions made in each appeal were also to be justified. If these assumptions are called into question then the justification of these appeals is the same.

The failure to recognize the essential identity of these two techniques is at least partly a product of the failure to appreciate that the destructive use of ordinary language requires commitment with respect to common sense epistemology. This is not true of the constructive reflection upon language, at least not in the same direct way. One deals with the actual circumstances, the other with conceivable circumstances.

This very blind spot is partly a testimony to Moore's early success. For when the so-called ordinary language movement was born, the philosophical world was already predominantly populated by realists, predisposed to accept common sense rather than to argue for it.

In the end I think it must be admitted that both of these appeals are little more than opening moves as far as the solution of philosophical problems is concerned. They can be used to put a certain pressure on a philosopher, to get him to see that something has gone wrong, hopefully, to convince him that he must either give up his thesis or re-express it. And it may be that calling the mistake a "misuse" has the advantage of indicating something about the nature of what has gone wrong. It

suggests that it is not simply false in the manner of yesterday's science. But this is its only advantage.

The urge to misuse language, or what comes to the same thing, the urge to say things that are false under conditions where it would usually be considered obvious that they are false, is symptomatic of deeper, more fundamental misconceptions. To show that a philosopher is misusing language is at best to *unearth* a philosophical problem not to solve it. It may be that the greatest weakness of the destructive use of ordinary language is not so much a matter of its legitimacy as of its limitations.

Nor would one wish to deny the very real chasm separating the work of Moore from that of the later Wittgenstein. Moore has said of himself that he was one whose interest in philosophy stemmed primarily from what other philosophers had said. What he tried to do was to show that what they said was wrong. It remained for Wittgenstein to broach the more insightful question as to why they should have wanted to say these things at all.

IV. MOORE'S PRINCIPLE

We have looked briefly at the respective end-games of two techniques. In fairness to Moore we should now like to suggest that it would be quite wrong to identify the essence of Moore's technique with his end-game. We have acknowledged that Moore can look excessively dogmatic if one focuses solely on this aspect of his method. His persuasiveness derives much more from a characteristic mode of argument. I refer to the way in which he was always turning the tables on his opponents, using their own arguments against them, hoisting them, as the expression goes, with their own petards.

A typical specimen of this Moorean maneuver is found in his argument against Hume:

If Hume's principles are true, then, I have admitted, I do not *know* now that this pencil—the material object—exists. It seems to me that, in fact, there really is no stronger and better argument than the following. I do know that this pencil exists . . . therefore, Hume's principles, one or both of them are false.⁶

⁶ *Some Main Problems of Philosophy* (London, 1953), pp. 119–120.

⁷ *Philosophical Papers* (London, 1959), p. 226.

⁸ *Ibid.*, p. 247.

⁹ *Ibid.*, p. 41.

¹⁰ *Ibid.*, p. 228.

¹¹ No doubt other formulations have been given. See R. Bambrough, "Principia Metaphysica," *Philosophy*, vol. 39 (1964), p. 103.

¹² This is roughly the informal equivalent of the relation between *modus ponens* and *modus tollens*.

¹³ "George Edward Moore," *ibid.*, p. 180, note 36.

He argues in a similar manner against Russell,⁷ the dream argument,⁸ and certain anticipated objections in his "Defense."⁹ The general character of his argument is clear in the following:

This after all, you know, really is a finger; there is no doubt about it; I know it, and you all know it. And I think we may safely challenge any philosopher to bring forward any argument in favor either of the proposition that we do not know it, or of the proposition that it is not true, which does not . . . rest upon some premiss which is, beyond comparison, less certain than is the proposition which it is designed to attack.¹⁰

What is important here is the role played by *relative* certainty.

The appeal to common sense is not as all-important as it first seems. What Moore has shown is that the arguments of the paradox propounders (if credible) are the instruments of their own destruction. In each case Moore makes use of a principle, which we might therefore appropriately call *Moore's Principle*.¹¹ It goes like this: *For every argument supporting a conclusion there is an equally credible counterargument that undermines the premiss.*¹² Thus in evaluating any argument it is always necessary to weigh the likelihood of the truth of the principles from which one is arguing against the likelihood of the falsity of the consequences derived from these principles.

It is in no small measure Moore's use of this "seesaw" principle that makes him effective. He did more than raise doubts about specific philosophical claims. He raised doubts about the nature of the conclusions that can plausibly be established by philosophical argument.

Perhaps we can illustrate this by calling attention to a fatal flaw in the account of Moore and the paradox propounders given by Malcolm. His interpretation rests on a principle which is often useful and might therefore seem above suspicion. He holds that "in order to tell what kind of thesis a philosopher is maintaining we have to consider the kind of support he offers for it."¹³ Relying on this principle Malcolm examines the arguments of the paradox propounders and concludes that their

theses are non-empirical in character. But Malcolm's principle requires us to assume that either the philosophers in question knew what sort of support was appropriate for their claims, or that even if they did not know, it so happened that the support they offered was in fact the sort required.

In this instance this is exactly what we can not assume. The paradox propounders were confused on just this point. They were oblivious to the intrinsic limits upon the factual discoveries that can be sustained by philosophical considerations. They

were caught up in the erroneous belief that arm-chair reflection and analysis could turn up startling and paradoxical new truths about the world; "Motion is impossible," "Time is unreal," "We only see our own brain," and so on. It is this fundamental misconception that Moore did much to discredit. By employing his principle and relying on common sense, he was able to meet the paradox propounders on their own ground, treating their claims as they saw them, empirical claims about the nature of our world.

Washington State University

Received May 24, 1968

VIII. THE NOTION OF HUMAN RIGHTS: A RECONSIDERATION

MEIRLYS OWENS

I

I am not so much interested in asking "Are there any human rights?" or "Which are the human rights?" as in asking: "In what context does one speak of a human right? How is the notion used in the various declarations of human rights or the rights of man?" In 1776 it was proclaimed that all men are created equal, and that they are all endowed by their creator with certain inalienable rights among which are the rights to life, liberty, and the pursuit of happiness. In 1789, to these inalienable rights were added the right to property, security, and resistance to oppression. The Universal Declaration of Human Rights proclaimed by the General Assembly of the United Nations in 1948 contains 30 articles which give a detailed enumeration of what a human being, a society, or a country may justly claim and fight for. But why should such an enumeration be important?

What is emphasized is the role which such a declaration has in serving as "a common standard of achievement for all peoples and all nations."¹ The "achievement" of any state or nation will then be measured in terms of how many of the declared rights have become "positive" rights which are recognized by the particular laws of particular states. Not only will diverse ambitions and activities be limited by the common measure, but they will also be unified and inspired by the end implicit in such a measure. Moreover, it is only this measure which can justify the interference (armed if necessary) of united nations in the affairs of other countries: so that injustice, suffering, wholesale massacre, and destruction may be prevented.

My task does not lie in trying to determine how effective such a measure has actually been. Rather, my concern is with the objections of those who, while fully appreciating the worthy aims of

the men who formed and advocated such a measure, find that very measure inadequate. What does it mean to talk of "inadequacy" here?

Surely it must be recognized that the declaration of human rights is rooted in a concern for other human beings. It is "a solemn protest against the brutal oppression, torture and assassination" associated with certain forms of government.² What had to be found was *the* measure of what is just and unjust, the measure which does not change with time and with conditions which may or may not prevail on this earth. So it was asked—just where does the difference lie between acting justly and acting unjustly toward other human beings? What prevents me from doing anything I please—imprison him, torture or kill him—to any human being?

"Legal restrictions" could form one answer, "social rules" another. Callicles would surely say: only my lack of sufficient nerve.³ But all these answers bypass considerations concerning the nature of an action and its effect on a human being: that the life of this particular person has been preserved is only incidental. If the law was robbed of its supremacy, or social rules lost their force, or circumstances arose in which I could kill without danger of punishment, then I would kill.

What kind of answer is immune to such relativity? In Dostoevsky's *Crime and Punishment*, Raskalnikov desperately attempts to justify his murder of the moneylender by calling attention to the good he did the world by ridding it of a mean nasty old louse. But Sonia's simple question puts an end to all such justification. She merely asks: "A human being—a louse?" And this whole way of talking cuts deeply into the difference between justice and injustice, good and evil in the lives of human beings. For it is connected with the sense of degradation that evil doing and unjust actions leave in a man's soul. To most men there is something

¹ See the Preamble to the Declaration of 1948.

² M. Moskowitz, *Human Rights and World Order* (London, Steven, 1959), p. 157.

³ See Plato's dialogue, *Gorgias*.

terrible and degrading about murder and all forms of injustice: "I can't do that to him. He is a human being, not an animal." Here the "can't" is not relative to physical possibility or impossibility, nor to what the law or society forbids. Neither is it a "can't" which is solely dependent on the particular relations between oneself and others where love and affection determine how I can and cannot act toward parents or friends, for example. But it is a "can't" which enters into one's relation with *any* human being, and which is inseparable from a respect for the individual who confronts one, nothing else, neither one's own interests, well-being, nor one's social standing being taken into account. It is this concern which is at the very center of every moral action; it is this respect which ought to be given to a human being because he is of intrinsic, infinite value, this value manifesting itself in the absolute rights which belong to him as a person. To be unjust is to ignore or to violate those rights. To be just is to respect them and to endeavor to create those conditions in which men may live their own lives, pursue happiness, and fulfill their destiny.

II

Why, therefore, the charge of "inadequacy"? Someone who has made such a charge is Simone Weil. Her analysis of human rights and human obligations in *The Need for Roots*⁴ is of particular importance since it was demanded by a real situation: by the urgent necessity of regenerating a France made moribund by war and German occupation. It is this analysis which leads Simone Weil to maintain that to speak of human rights as absolute principles is "to tumble into a confusion of language and ideas which is largely responsible for the present political and social confusion." But why isn't it possible to use such terms as "unconditional" or "absolute" when talking of human or minimal rights?

We may speak of the rights of prime ministers, of judges, of chaplains, or of parents. These rights are related to how things are or ought to be within social institutions or forms of social organization. They may change with time and circumstances: some may disappear as different forms of social organization develop, or some may change in character as institutions change. As various the rules connected with innumerable institutions and ways of acting within a society, so as various are the rights which men have. And Simone Weil recog-

nizes this. But what she also wants to emphasize is that the notion of rights is *inseparable* from the notions of existence and reality, and that rights are *always* found to be related to certain conditions. That is, it is pointless to speak of the rights which all men possess by virtue of their being human, *in vacuo*, or without considering the conditions that prevail and how the rights can be honored or granted. Indeed Article 29 of the Declaration of 1948 itself makes it quite clear that the rights of each individual can be exercised only within the limits of "the just requirements of morality, public order and the general welfare in a democratic society" where considerations of utility and expediency will also play a part. So it may be said that the "absolute value" of the human being is made subject to laws and institutions, to what is useful or possible.

But this is an observation which has been made by other critics of the notion of human rights or the rights of man, for example Edmund Burke who maintained that the situations in which men may find themselves are too great for a prior enumeration of all the rights which men might reasonably claim in each of these situations. So where does the significance of Simone Weil's criticism lie? It lies in its connection with the distinction she draws between the notion of rights and that of obligations. For what she maintains is that the notion of obligation comes *before* that of rights which is subordinate and relative to the former. But this may be difficult to understand. What kind of "priority" is this?

Perhaps it is important to note first of all that if one says that the notion of obligation is prior to that of rights, this is not the obligation to *political* obedience. We find that rights are generally asserted in opposition to an existing, or newly overthrown regime (as was the case in 1776 and 1789, and also with the English *Bill of Rights*). Or it might be the case that rights are asserted to limit the sphere of government action. The Trade Union Movement worked to obtain the recognition of certain rights in regard to workers' organizations and collective bargaining, etc. What is apparent is that reactionary, conservative, or totalitarian politicians dislike any agitation by trade unionists or others for the recognition of rights. Stanley Baldwin, for example, in the 1930's told a meeting of students: "Put your *duty* first, and think about your rights afterwards." One of the Soviet Russian politicians, Vishinsky, once told his audience: "You have no rights, only

⁴ London, Routledge & Kegan Paul, 1952. This discussion which follows owes a great deal to the insights of that analysis.

duties." These are merely ways of saying: "Don't think of revolt."⁵ Simone Weil would not want to say this. On the contrary, she emphasizes that we owe respect to the state, or to an institution or to a form of collective life, not for itself, but because, and only because, it is food for a certain number of human souls. If it fails to provide that food, then it should either be improved or destroyed. But one has direct duties of an eternal nature *only* to the human being for the sole reason that he *is* a human being, without any other condition requiring to be fulfilled, and even without any recognition of such obligation of the part of the individual concerned.⁶

But why should it be said that the notion of obligation is prior to that of rights? Isn't it the correlativity of the notions which is so often emphasized? A.I. Melden, for example, says that while on the one hand the right of the parents to respect and consideration implies a correlative obligation on the part of the son to give them that respect and consideration, on the other hand, the obligation on the part of the son to honor his parents implies the correlative right of the parents to such consideration.⁷ But the "implication" here has significance only within what Melden calls "a common moral understanding" where obligations *are* recognized and fulfilled, and rights *are* respected. The correlativity of the notions is important when we consider how we come to know what our obligations are, or what an obligation is by participating within ways of living where common understanding between human beings gives life and substance to the notions of rights and obligations. Melden goes on to show that even within this common moral understanding, there are significant differences between the ways we talk about rights and about obligations, which show that the relation of a right to what Simone Weil calls "the absolute good," is very different from that of an obligation to "the absolute good."

But when one emphasizes the *priority* of the notion of obligations to that of rights, one is not saying anything about how different rights and obligations work within a society, neither is one concerned to elucidate what is involved in a common understanding, but one is concerned with the *nature* of a right, and the *nature* of an obligation and their

respective relation to understanding that there is an absolute and unconditional difference between good and evil: one wants to draw attention to the relation between obligations and moral actions.

But how *is* the notion of obligations prior to that of rights? The answer has to do with the way in which we do speak of obligations as being absolute and unconditional. It could, of course, be asked: "But aren't obligations as much related to various conditions, circumstances, etc., as rights are?" Simone Weil, however, never denies that when an obligation descends to "the realm of facts," it involves taking into account what kind of actions are necessary in different situations, but she does deny that any combination of circumstances could ever cancel, for example, the obligation to remedy all the privations of body and soul which may destroy or damage the earthly life of any human being whatsoever.

The "priority" is also important when one considers the kind of relation which holds between the recognition and the fulfilling of an obligation. It is a relation which Simone Weil describes by saying that while the effective exercise of a right springs not from the individual who possesses it, but from other men who consider themselves as being under an obligation toward him, the very recognition of an obligation makes it effectual.⁸ Although the recognition of an obligation does allow for the failure to fulfill it, that failure would necessarily result in remorse, or a deep feeling of guilt, or in an awareness of one's moral weakness. The necessity which belongs to an obligation is such that if the obligation is not fulfilled, the degradation which belongs to evil doing or injustice will penetrate one's whole soul. That is why, when someone is faced by conflicting obligations, the inevitable sacrifice of one arouses deep remorse, even though it was the fulfilling of another obligation which demanded such a sacrifice.⁹ But the failure to satisfy a right may not go together with any such feeling of remorse. On the contrary, there may be circumstances where it is morally praiseworthy *not* to satisfy a right, and not to expect a right *to* be satisfied; or to forego a right. This can *never* be the case where an obligation is concerned. That is why Melden emphasizes in *Rights and Right Conduct* that

⁵ I am indebted to Mr. Rush Rhees for drawing my attention to the importance of this distinction, and for Baldwin's and Vishinsky's remarks.

⁶ *The Need for Roots*, *op. cit.*, p. 5.

⁷ *Rights and Right Conduct*, Sects. I-IV, (Oxford, Blackwell, 1959).

⁸ *The Need for Roots*, *op. cit.*, p. 3.

⁹ See the discussion by D. Z. Phillips and H. S. Price on "Remorse Without Repudiation" in *Analysis*, vol. 28 (1967), pp. 18-20.

it does not make sense to talk of exceptions where ought-statements are concerned.

III

But given this distinction between the nature of a right and that of an obligation, couldn't a case be made for the uniqueness of the appeal to human rights? Melden, for example, wants to argue that one of the "significance-conditions" of such an appeal is that the one who makes it wants to effect certain social changes in order to realize the aspirations of an oppressed people; that is, an appeal to human rights is meaningless unless it goes with a recognition of an obligation to act on behalf of such a people.¹⁰ Doesn't this mean, however, that the "priority" of the notion of obligations has little significance where an appeal to *human* rights is in question?

But in what context would such an appeal be made? Surely, as Melden observes, in that context where as a result of military subjection and exploitation, the family and local community life of the natives, their farms and their villages, are disrupted, where human beings are left to eke out a wretched existence. "Clearly the human rights of these natives to a variety of things are violated."¹¹ So that an appeal to human rights is made when what is at stake is the *establishment* of a community in which moral regard may flourish or in which moral discourse may have application. Perhaps Simone Weil wouldn't disagree with this. But what does concern her in *The Need for Roots* are the special problems connected with "establishing a community." She maintains that when the primary concern is that of finding a way of inspiring a people to effective and creative action after a period of lethargy and disintegration, then it is the recognition of certain obligations rather than the recognition of rights which is important.

How is the restoration or renewal of a community or a society to be effected? How is it possible to breathe an inspiration into a people uprooted by war and occupation? Not *any* kind of method will do—or else the inspiration itself will be destroyed. Some awareness of this problem is shown in the Preamble to the Declaration of 1948 when it emphasizes the need for teaching and education to promote respect for the rights and freedoms of the human person. But nothing is said about the

special difficulties connected with such an education. How are people to be so inspired by what is beneficial or what is obligatory, so that this inspiration "passes into the muscles and emerges in the form of actions"? One thing is clear: There is no technique for doing this; no set of rules which may be applied and which guarantee results. In his discussion on the rights of man, Maritain states that it is possible to create respect for the rights of others by taming the irrational to reason, and by developing the moral virtues by constant training: that one comes to love the good by being trained to do good actions. But surely a love of the good is the source of all moral actions and not their outcome. You may *teach* someone that all men have certain minimal rights, without the exercise of which they become little more than animals, and that is all you *can* do. You can't instruct a person to have a regard for those rights, something other than instruction is necessary. In the *Phaedrus*, Plato says that beauty is the only sustenance that can make the wings of the soul grow, and what this means is that a recognition of beauty in some form or other is the source of all regard, whether it be for human beings, for laws and institutions, or for certain actions. Again in the *Symposium* it is seen that a love of the good is inseparable from a recognition of the beauty of laws, of sciences, of institutions, of customs and traditions, and of amity between men.

All this is closely connected with what Simone Weil wants to say about the importance of inspiration in relation to the actions necessary for the regeneration of a society. We all know that the directing of human beings is effectively and easily done by the use of threats and promises, or by suggestion; but the energy which is infused into people by such means is an alien one. The actions done purely out of a regard for another human being, or regard for an institution or a form of collective life which provides food for a number of souls, demand an energy which is not imposed nor created by something which is different in kind from the actions that are done. That is why Simone Weil speaks of the role that friends, or natural guides have of inspiring and of infusing energy into those who respect their words. Where friendship, admiration, sympathy, and respect are present, so also is a readiness to give attention to the words that are uttered—words which become "food" or "energy" to the one who listens. When the inspiration of

¹⁰ See his paper on "The Concept of Universal Human Rights" in the *American Philosophical Association*, vol. I (1952), pp. 167–188.

¹¹ *Ibid.*, p. 185.

a whole people is at stake, this role must be taken over by a government or a movement whose authority must not be based on physical power which can be destroyed by defeat, nor on glory which can be wiped out in shame; but must proceed on an elevated plane of thought in keeping with the plight of a people and on a spiritual tradition graven on the hearts of all peoples.¹² Not only must a movement know which words will find an echo in the hearts of the people who are to be inspired, and not only must it also examine and weigh with extreme care all the methods of action which form part of the process of the growing of roots, but it must also create and sustain organizations "which crystallize and seize upon the words launched officially, translate their inspiration into different words entirely of its own, realize them in co-ordinated actions for which it offers an ever increasing efficacy."¹³

IV

But what are these "words"? In an essay on the *Human Personality*,¹⁴ she observes that for the aspirations of the afflicted, if we wish to be sure of using the right words, all that is necessary is to confine ourselves to those words and phrases which always, everywhere, in all circumstances express only the good. That is why the language of obligations rather than the language of rights must be used. Only obligations belong to "the realm of pure good." "To possess a right implies the possibility of making good or bad use of it; therefore rights are alien to good. On the other hand, it is always and everywhere good to fulfill an obligation. Truth, beauty, justice, compassion are always and everywhere good." These then, are words which for Simone Weil possess in themselves, when they are properly used, a virtue which illumines and lifts toward the good.

That is why in *The Need for Roots*, the notion of rights is replaced by the notion of obligations related to needs. The very notion of a right suggests a claim or demand, a resilience, a certain status. "It has a commercial flavour, essentially evocative of legal claims and arguments. Rights are always asserted in a tone of contention; and when this tone

is adopted, it must rely upon force in the background or else it will be laughed at."¹⁵ This is why Kierkegaard¹⁶ talks of selfishness and self-love when he talks of the concern for rights, whether they be my rights or the rights of all men. As Burke pointed out, what are called "the rights of men" can so easily become "allies of the flesh"—pretexts behind which "pride, ambition, avarice, revenge, lust, sedition, hypocrisy, ungoverned zeal and all the train of disorderly appetites" are hidden under the "specious appearance of a real good."¹⁷ The notion of a need, however, goes together with destitution, weakness, poverty, suffering, or affliction; there is something terrible and degrading in ignoring or denying a need. And it is this which is emphasized when Simone Weil says that there exists only a single, eternal, unconditional obligation toward each human being, that is, respect. For this obligation is only performed if that respect is effectively expressed in a real, not a fictitious, way; and this can only be done through the medium of man's earthly needs.¹⁸

For Simone Weil, the most obvious need is that of hunger. "Everyone looks on progress as being, in the first place, a transition to a state of human society in which people will not suffer from hunger . . . [no] man is innocent if possessing food himself in abundance and finding someone on his door-step three parts dead from hunger, he brushes past without giving him anything." Someone may not agree that this is the most obvious of all obligations: it may be said that the degradation which belongs to taking the life of a man is even more terrible than that of letting someone go hungry. Be that as it may, what is important here is the way Simone Weil uses this obligation as a model on which to draw up the list of eternal duties to each human being: the list comprised of those obligations which correspond to human needs which are vital, or analogous to hunger.

But then the difficulty is to say *which* needs are vital. Simone Weil would not want to say that the presence of a need always imposes an obligation to satisfy that need. On the contrary, there are some needs (for example, the need for social prestige—which must be clearly distinguished from the need for honor, however difficult it is to make that

¹² *The Need for Roots*, *op. cit.*, p. 189.

¹³ *Ibid.*, p. 205.

¹⁴ *Selected Essays*, ed. by Richard Rees (Oxford, Oxford University Press, 1962), p. 24.

¹⁵ *Ibid.*, p. 18.

¹⁶ See *Purity of Heart*, tr. by Douglas Steere (New York, Harper and Brothers, 1948), p. 96.

¹⁷ *Works III*, 418.

¹⁸ *The Need for Roots*, *op. cit.*, p. 6.

distinction perhaps) which it would be wrong to satisfy. But then she would want to distinguish between *needs* and whims, desires, fancies and vices, or between what is fundamental and fortuitous. That is why she talks of the needs of the body and the needs of the *soul*, rather than the needs of the human being. What are called "the needs of the soul" here may be misleading until one comes to understand that these needs are a necessary condition of the life of men here and now. The point of emphasizing this may be seen when one considers how someone connected with the Spanish Inquisition may have justified its purges by death and imprisonment by maintaining that the soul needed to be purified from heresy and sin. The needs of the soul which Simone Weil is concerned about are those which if ignored or denied result in a "state more or less resembling death, more or less akin to a purely vegetative existence": they are earthly needs.

But there is another special reason why emphasis is put on the needs of the soul and body rather than on the rights of the human being. It is connected with what is said about the need for order in the face of the variety of obligations which belong to a network of social relationships. Still employing the analogy of hunger and sustenance, Simone Weil notes that (a) needs are as limited as are the foods which satisfy them: to go beyond a certain limit once the hunger has been satisfied is to result in satiety; so the need for food on the part of the soul; (b) needs are balanced against each other. "Man requires food, but also an interval between his meals; he requires warmth and coolness, rest and exercise. Likewise in the case of the soul's needs."¹⁹ That is why the needs of the soul are, for the most part, listed in pairs of opposites which balance and complete one another: there is a need of liberty *and* obedience, of equality *and* hierarchy, of risk *and* security, of private *and* collective property.

The analogy she takes is an important one. Truth, obedience, liberty, honor, responsibility, equality, punishment, and freedom of opinion are *foods* of the soul—when this nourishment is lacking, so the soul is victim to "starvation" and "asphyxiation." But what are just as important as the sustenance are the ways through which the soul is sustained. That is why the analogy of rootedness and uprootedness is made central in the discussion: the need for roots is bound up with the whole of man's moral, intellectual, and spiritual life.

Uprootedness can have the gravest consequences upon the life of an individual or on that of a country. It is by means of his roots that a human being can make "contact with the universe"—by virtue of "his real, active and natural participation in the life of a community which preserves in living shape certain particular treasures of the past and certain particular expectations for the future."²⁰ It is through a man's country, his language, his culture and traditions, the place where he works, his neighborhood, that sustenance is given him. To be uprooted means that even the possibility of understanding his own life is withdrawn; there is a waning of that deep regard for those values which would give his life meaning, direction, and solidarity. So the past is impoverished, and the future devoid of hope and aspiration. Edith Wharton, who has shown how this rootlessness can disintegrate and destroy the lives of human beings, observes that in whatever form a slowly accumulated past lives in the blood, it has the power of broadening and deepening individual existence, "of attaching it by mysterious links of kinship to all the mighty sum of human striving." For Simone Weil destruction of the past is one of the greatest of all crimes. "The future brings us nothing, gives us nothing; it is we who in order to build it have to give it everything, our very life. But to be able to give, one has to possess; and we possess no other life, no other living sap, than the treasures stored up from the past and digested, assimilated and created afresh by us."²¹ To identify love of the past with a reactionary political attitude is to fail to understand that the source of the former is very different from that of the latter. To emphasize the importance of rootedness is not to rule out, nor to belittle the importance of examination and criticism of (and even revolt against) those conditions in which a human being may find himself. For not only would the intelligibility of the criticism be inseparable from the existence of those conditions, but the very criticisms would indicate a recognition of the importance of these conditions upon the lives of human beings.

Much more could be said in connection with this vital need for roots: the account just given can be little more than a general outline of what Simone Weil says in *The Need for Roots*. If it appears unsatisfactory, it is because I have failed to do justice with the detailed suggestions she makes

¹⁹ *Ibid.*, p. 12.

²⁰ *Ibid.*, p. 41.

²¹ *Ibid.*, p. 47.

IX. ACTIONS AS UNIVERSALS: AN INQUIRY INTO THE METAPHYSICS OF ACTION

CHARLES LANDESMAN

I

AN action is whatever anyone does. Are actions to be classified as universals or particulars or perhaps both? The traditional distinction between universals and particulars is reasonably clear in certain kinds of cases. The color red is repeatable and hence a universal because many different objects can be or are red. A red object is an instance of the color, and the object fails to be a universal because nothing is an instance of it. Qualities such as colors and shapes are repeatable and for that reason are entitled to be called universals.¹ Words are repeatable because the same word can be written or spoken on different occasions. If we count the number of words appearing in a copy of a book, we shall obtain different results if we count repetitions as being the same or different words. If we count them as being the same, we are counting types, otherwise tokens. Though word types are repeatable, they are not so in quite the same way as qualities. A token of the word "red" is a word just as its type. But an instance of the color red is not a quality. The relation *being a token of* is different from the relation of *being an instance of*.

If there is a sense of "action" in which two different persons can perform the same action, and another sense in which they cannot, then actions in both the repeatable and nonrepeatable senses are actions just as words in the repeatable and nonrepeatable senses are words. It seems, therefore, that if we begin by considering the question whether there is a universal-particular distinction to be

made among actions, the relevant contrast to consider is the one between types and tokens. Consider the example of a certain person's raising his arm. Let "(raising his arm)" designate what he did as something repeatable, as something someone else can do as well, and let "((raising his arm))" designate what he did as something non-repeatable. If a distinction between ((raising his arm)) and (raising his arm) can be made out, then the relation holding between the two is more similar to *being a token of* than it is to *being an instance of*. Thus the following view would appear to be a perfectly natural one to adopt: "Any token event *E*, and any token action *A*, are by definition particulars It is logically impossible that two persons should do the same *token* action."² Just as it is true that the same word can be written more than once if by "word" we mean type and that the same word can be written only once if by "word" we mean token, so, it is alleged, though the same action can be performed more than once, yet there is a sense of "action" analogous to that of "token" in which the same action can be performed only once. On this view, the distinction between universals and particulars can be drawn within actions just as it can be drawn within words. It is this claim that I shall dispute in the next section.

II

It appears that the distinction between word types and word tokens can be generated by adopting different methods of counting words.³ The

¹ If there are any attributes at all, then there are attributes such that it is logically impossible for more than one thing to be an instance of them, e.g., the attribute of being the only satellite of the earth. While these are not, strictly speaking, universals because they are not repeatables, they can be considered to be universals in a derivative sense because they resemble things that are universals by being attributes.

² B. A. O. Williams, "Personal Identity and Individuation" in *Essays in Philosophical Psychology*, ed. by D. F. Gustafson (New York, 1964), p. 328.

³ I am assuming that this widely employed distinction is clear in order to state the view I am about to criticize. That this assumption is doubtful can be gathered from the following example: Suppose in a copy of a book which had a printing of 1,000 we count 10,000 "types" and 20,000 "tokens." Yet each of these "tokens" reappears in each of the other copies. That is the 20,000 entities that were counted each recur 1,000 times and are not really tokens. If the term "token" is reserved for the specific nonrepeatable sign, we must distinguish between types of various levels, since something which is a repetition of a type

problem is to determine whether an analogous distinction can be generated for actions. If two different persons, Smith and Jones, each raises his own arm, then they both did the same thing. But if Smith raised his right and Jones his left arm, then what they did was not exactly the same. Suppose, however, that both of them raised their own right arms. Then didn't they do the same thing? Not exactly, since Smith did it slowly and Jones quickly. But if they both did it slowly, then they both did the same thing. What we are doing here is taking an action "raising his arm" and providing a series of more detailed descriptions of it: "raising his right arm," "raising his right arm slowly," etc. In each case, the question whether or not Smith and Jones did the same thing depends on whether or not the same details apply to what they did. And in each case, it was possible for them to have been doing the same thing. We did not arrive at a point at which it becomes logically or definitionally impossible for them to perform the same action. Merely by adding more details to a description of what a person does, we do not generate an action-token, an action in the nonrepeatable sense. As far as we have gone in this analysis, an action is like a color; providing a more detailed or specific description of a color—for example, red, dark red, crimson, etc.—still leaves us with a universal. The relation between "raising his right arm" and "raising his arm" is more like the relation of a determinate to its determinable—such as crimson to red or square to rectangular—than it is like the relation of a token to its type.

The advocate of the type-token distinction for actions will now step forward and argue as follows: "The sense of 'action' in which what Smith does is nonrepeatable is that what Smith does is done by Smith. Although the action (raising his arm) is a repeatable entity because more than one person can raise his own arm, nevertheless the action ((Smith's raising his arm)) must be nonrepeatable because only Smith can do that." But the argument is still incomplete, for even if it is conceded that only Smith can do that, yet he can do it more than once. It is necessary to augment the argument as follows: "To guarantee nonrepeatability, reference must be made not only to the person but also the time of the action's occurrence. It is ((Smith's raising his

arm at 10:03 p.m., January 3, 1792)) which is the nonrepeatable entity we are after. Not even Smith could do that again."

To assess this line of argument, let us change the subject temporarily and reflect on the nature of colors rather than actions. Suppose someone were to argue in a similar vein that a type-token distinction can be made for colors as well as for words. The color red, for example, is a universal of which individual red things are instances, but there is also the particular nonrepeatable redness of this red thing. But, we might well ask, how can we distinguish between the color red and the particular redness of this red thing? We are not, remember, concerned with the distinction between the color red and the particular shade of red that this thing has, because the distinction between a color or a shape determinable and one of its determinate forms is not a version of the universal-particular contrast. No matter how specific and determinate a shade of color we mention, what we mention is repeatable; another red object may have just that shade. Perhaps the advocate of the type-token distinction will concede this point. "I agree that the particular nonrepeatable redness of this red thing is not to be identified with the specific shade of red that this thing has. What we are referring to is the particular redness which this thing has at this time. That surely is nonrepeatable."

The difficulty with this line of reasoning is that it smacks of magic, of producing something from nothing, or more precisely, of generating a new class of objects merely by describing an old class in a new way. The magic formula is this: to generate an entity y from any given entity x simply select something z to which x has some relation R , and let y be the x having R to z at time t . Thus given some specific shade of red x , we are said to generate a color token y of that shade by taking as z the object having that color and R the relation between z and x when z is an instance of x . The color token is allegedly specified as the color which that object has at just that time. The fallacy in the argument is that although " x " ("the color red") and "the x having R to z at t " ("the color which that object has at that time") are different phrases, we do not as yet have a reason to think that they designate different things. The magic formula is a way of

need not always be a token. The distinction we are criticizing has been formulated without using the token-type terminology by von Wright: "The word 'act' . . . is used ambiguously in ordinary language. It is sometimes used for what might be called act-qualifying properties, e.g., theft. But it is also used for the individual cases which fall under these properties, e.g., the individual thefts . . . The individual cases that fall under theft, murder, smoking, etc., we shall call act-individuals." ("Deontic Logic," reprinted in *Contemporary Readings in Logical Theory*, ed. by Irving Copi and James Gould, [New York, 1967], pp. 303-304.)

generating a new description out of an old one, but it does not guarantee that a new object can be generated from the old one. Why can't some repeatable shade of red be identical with, be the very same thing as the particular redness of this thing at this time? So far, no argument has been offered that they must be different, and we now realize after modern discussions of identity that differences in descriptive phrases do not necessarily mark a difference in the things described. Moreover, not only is there no reason in favor of drawing a type-token distinction for colors but there are two good reasons against it. The first is that when a person sees that a particular object is red there is no further observation he can make to discriminate between a repeatable (redness) and a particular ((the redness of this object at this time)). Even if he were to improve his visual acuity, by wearing spectacles for instance, there would still be no distinction for him to notice. The type-token distinction for colors is not observable in any obvious way. Second, all things being equal, considerations of metaphysical simplicity militate against introducing an additional category of entities—color tokens—in the absence of strong positive reasons.

Similar considerations apply to actions. Merely by expanding the phrase "raising his arm" to "Smith's raising his arm at 10:03 p.m., January 3, 1792," we have no guarantee that a new category of entity has been distinguished. Merely asserting emphatically that it is the action done *by Smith at 10:03 p.m., January 3, 1792* that we are talking about does not by itself preclude someone else from doing at a different time just what Smith did at that time. In short, there is nothing about the procedure invoked to generate a type-token distinction for actions which guarantees that if we begin by acknowledging actions in the repeatable sense the procedure will lead us to an action in the non-repeatable sense.

The thesis that there is a type-token distinction to be drawn for actions, besides lacking an adequate positive justification, has certain difficulties of its own. In the first place, when we use Smith's name in describing his action as in "Smith's raising his arm" it appears that we are not merely identifying what action was done but also who did it. Similarly, when we state the time of the action we are not only identifying what the action was but when it was performed. It appears, then, that the description "raising his arm" is not an incomplete designation of what Smith did which can be completed by mentioning that Smith did it at 10:03 p.m., January

3, 1792. It is, rather, a complete designation of an action which fails to mention who performed it and when it was performed. It would be absurd to claim that the phrase "of this egg" drawn from the description "the oval shape of this egg" contributes to indicating what the shape is rather than what has the shape; it is similarly absurd to claim that the phrase "of this couch" drawn from "the orange color of this couch" contributes to telling us what the color is as distinguished from what has the color. It is similarly absurd to say that the verbal fragment "Smith's . . . at 10:03 p.m., January 3, 1792" drawn from "Smith's raising his arm at 10:03 p.m., January 3, 1792" contributes to telling us what action was performed instead of who did it when.

In the second place, suppose, contrary to our view, that "Smith's raising his right arm at 10:03 p.m., January 3, 1792" does designate a non-repeatable action. We can then ask: Who performed that action? Obviously it was either Smith or someone else. But it couldn't be anyone else, so it must have been Smith. But what was it that he was supposed to have done? He raised his arm, that is for sure. But his raising his arm is not the action we want since *that* could be done by someone else. After all, Jones can raise his arm as well. The non-repeatable action Smith performed was Smith's raising his arm at 10:03 p.m., January 3, 1792. Although, no doubt, Smith can raise his arm, can Smith really do the act of Smith's raising his arm? And when did he do it? The action of raising his arm, we now know, is something he did in 1792. But the action whose time of occurrence we are now asking for is not "Smith's raising his arm" but "Smith's raising his arm in 1792." Certain questions that we sensibly raise about human actions such as who did what and when was it done have a clear sense when asked of actions conceived as universals, but appear to lack a sense when raised of action tokens.

Finally, it follows from the view being criticized that the sentence "Smith raised his arm at 10:03 p.m., January 3, 1792" is ambiguous. It can be interpreted to mean either that a certain action (type) was performed by Smith at the time indicated or that a certain action (token) exists. However, the truth conditions of both interpretations are exactly the same; whatever verifies the one also verifies the other. Moreover, none of the words in the sentence are particularly ambiguous. It is not a sentence like "Her chest is large" which, so to speak, cries out for different interpretations.

Thus we have here a case of a theory requiring that each of the members of the set of action sentences is ambiguous where there is every reason to think, independently of the theory, that the alleged ambiguity is fictitious. Here is a further reason to judge that the theory is false. Just try to utter the sentence, first meaning to assert the first interpretation and then the second. What difference is there that is discernible?

III

Given that actions are universals rather than particulars, that they are repeatable entities, that different persons can perform literally the same action, there still remains the problem of elucidating the type of universal involved. How do actions compare with qualities, relations, species, and so forth? The following way of answering this question involves several rather familiar ideas of modern logic and semantics. Consider any meaningful declarative sentence containing one or more singular terms. A sentence frame is what is left when at least one of the singular terms is erased and a variable placed in the empty position. A predicate is that part of the sentence frame other than the variable(s). An attribute is that entity designated by a predicate.⁴

Thus, from the sentence "Socrates is a man" we derive the sentence frame " x is a man" which contains the one-place predicate "is a man" which designates the attribute *being a man*. By parity of reasoning, from the sentence "John runs" we construct the sentence frame " x runs" which contains the predicate "runs" which designates the attribute *running*. We can think of actions as a type of attribute, and this is a further justification for taking them to be universals. Some actions belong to that class of attributes known as relations. In the sentence frame " x killed y " the predicate "killed" designates the action *killing*. This is a two-termed attribute or relation. Just as we say that Socrates is an instance of *being a man* and a red object an

instance of *redness*; so now we can say that John is an instance of *running* and Brutus and Caesar jointly and in that order instances of *killing*. This analysis reveals a linguistic similarity among actions, qualities, species, and relations because it shows that words designating each of these can occupy predicative positions in sentences. Of course there are important differences as well. Unlike *being red*, *being a man*, and *being next to something*, an action is what someone does. A person is not merely an instance of his actions, but in performing them he brings it about that he is an instance.

We are now in a position to draw a distinction relevant to the argument in the preceding section. In the statement (a) "Smith raised his arm at 10:03 p.m., January 3, 1792" the action that Smith is said to have performed, namely raising his arm, is a universal. But its being performed by that person at that time certainly is not the sort of thing concerning which it makes sense to say that it is repeatable. Let us call this sort of thing a *fact* which can be represented either by (b) "the fact that Smith raised his arm at 10:03 p.m., January 3, 1792" or by (c) "Smith's having raised his arm at 10:03 p.m., January 3, 1792." The truth of (a) entails the existence of the fact represented in (b) and (c), and their existence entails (a).

Let us call the facts which "contain" actions in the way that (b) and (c) contain Smith's action *action exemplifying facts*. We distinguish the relation between Smith and his action which we mark by saying that Smith is an *instance* of the action from the relation between the action exemplifying fact and the action which we mark by saying that the fact *exemplifies* the action. If we focus on formulation (c) rather than (b) we might be more inclined to say that what (c) represents is an event rather than a fact. We shall not yield to that inclination; instead we shall use the term "event" to designate a class of universals or repeatables. The predicate of the sentence "The dynamite exploded" designates an event of which the dynamite was an instance. Facts which "contain" events in the way

⁴ I do not intend the expression " x designates y " to mean the same as " x means y " or " y is the meaning of x " or " x names y ." In "John walks" the predicate "walks" designates the action of walking whereas "walking" in "Walking is good for health" names the very same action. When a speaker utters the sentence "The book is red" to make the statement that the book is red, he is doing something which may be described as ascribing a color to the book, the color so ascribed being red, or describing the book as being red. Ascribing the color red to some object or describing it as being red is the standard thing people do with singular affirmative subject-predicate sentences whose predicate is "is red." This standard ascription is possible because of an association that has come into existence between the predicate and the color red. In general, people do things with predicates in virtue of the associations they acquire with attributes; they can do different things with different predicates in virtue of their associations with different attributes. I am using "designation" as the name of that association between a predicate and an attribute. The meaning of the word "red" is not unaffected by the fact that "is red" designates red, but it is not to be identified with that fact.

(b) contains an action we can call *event exemplifying facts*. Actions are a species of events, and events a species of attributes. I shall not in this context discuss how to demarcate actions from other events, and events in general from other attributes. Nothing in the argument is intended to imply that the distinctions here delineated are perfectly precise or exhaustive.

The distinction between actions and action exemplifying facts may very well have been mistaken for a type-token distinction by those who argued for the latter. Since the same action may be exemplified by numerous action exemplifying facts, there is some analogy to types and tokens. The error involved in drawing the type-token distinction is the claim that an action must be different from an action *qua* exemplified in some action exemplifying fact; as we have seen in the previous section this claim is gratuitous.

IV

Occasionally in the affairs of ordinary life we wish to know how many actions a person performed through some period of time. We raise questions about the identity of actions—Did Mr. So and So do three different things or was it the same thing referred to in three different ways? In a much discussed article, Donald Davidson has written:

I flip the switch, turn on the light, and illuminate the room. Unbeknownst to me I also alert a prowler to the fact that I am home. Here I do not do four things, but only one, of which four descriptions have been given.⁵

The four descriptions are "flip the switch," "turn on the light," "illuminate the room," and "alert the prowler." Davidson is claiming that the thing of which these descriptions are descriptions of is the action performed. We may surmise that he is here thinking of actions as particulars which are themselves instances of attributes or subjects of possible descriptions. However, each of these four descriptions are predicates, and the only clear sense in which a predicate can be a description of something is in which the thing described is referred to by the subject term, in this case by the term 'I'. On our view, it is the subject not the action which

is being described. If such predicates as "has red hair," "is six feet tall," and "is thirty years old" can be used as descriptions of persons, certainly so can "flip the switch" and "turn on the light."

The question: "How many actions did the person perform?" reduces, on our view, to the question: "How many attributes are designated by the four descriptions?" Now the topic of attribute identity is one of the more problematic areas of semantical theory, and I have no intention of dealing with it in any systematic way here. There is, however, a principle which if it could be shown to be acceptable would be a partial solution to the problem and which can help us here. It states a sufficient condition for the nonidentity of attributes and may be given a rough formulation as follows: (Q) "Predicates which are logically independent of one another designate different attributes."⁶ This is not a novel principle; it has been stated or implied in other writings.⁷ However, there have been counterexamples proposed which are alleged to refute it.

Consider, for example, the following pair of statements:

- (a) Jones went to the beach yesterday.
- (b) Jones did what Smith did yesterday.

If we assume that Smith also went to the beach yesterday, one may argue that the predicate of (a) "went to the beach yesterday" designates the very same action as the predicate of (b) "did what Smith did yesterday" even though they are obviously logically independent. The action that Smith performed was to go to the beach; the intent of (b) is to ascribe that very same action to Jones; thus the predicate of (b) must, it is argued, designate the very same action as that designated by the predicate of (a).

There are two ways of dealing with this alleged counterexample. One way, whose use I favor in this case, is simply to deny that it is a counterexample; that is, to deny that the two predicates designate the very same action. The predicate "did what Smith did yesterday" does not ascribe to Jones a particular action; what it can be used to say in (b) is that there is some action Jones performed which is identical to an action of Smith's. Someone who knows that Smith went to the beach

⁵ "Actions, Reasons, and Causes," *The Journal of Philosophy*, vol. 60 (1963), p. 686.

⁶ Predicate *A* is logically independent of predicate *B* if and only if the statement that some particular thing is an *A* neither entails nor is entailed by the statement that that thing is a *B*.

⁷ It is entailed by a plausible interpretation of Carnap's convention "Two predicators have the same intension if and only if they are *L*-equivalent." See his *Meaning and Necessity* (Chicago, 1960), p. 19. For a similar principle see C. I. Lewis, *An Analysis of Knowledge and Valuation* (La Salle, 1950), p. 64.

yesterday may gather from (b) that Jones also went to the beach. But not everything that one can gather from a sentence using independent information is part of the meaning of the sentence or of any of its constituents.

There is an alternative way out for those who are convinced intuitively that the two predicates do designate the same action and that there are consequently actual counterexamples to (Q). For even if the predicate of (b) does designate an action, it does so without saying what the action is. To make this clear, consider the description "the last proposition Plato thought of before he died" (D1). Suppose the last proposition he thought of is the one also named by the description "the proposition that two plus two equals four" (D2). Even though both descriptions name the same proposition (D2) tells us what the proposition is whereas (D1) does not. We can see this by noting that (D2) does but (D1) does not provide an answer to the question "What was the last proposition Plato thought of before he died?" Let us say that (D2) *informatively* denotes the proposition that two plus two equals four whereas (D1) non-informatively denotes the same. Similarly (a) does but (b) does not answer the question "What did Jones do yesterday?" Thus if the two predicates do designate the same action, that of (a) does so informatively whereas that of (b) does not. That there is a distinction between informative and non-informative denotation and designation is obvious even though it is not clear how to formulate the criteria of the distinction. It is possible to save the principle (Q) by reformulating it as follows: "Predicates which are logically independent of each other and which are informatively designative designate different attributes."⁸

It is possible to defend principle (Q) against certain types of arguments. Is there anything to be said in its favor? If we recur to Davidson's claim that the four descriptions were of only one action, we can now see that (Q) enables us to provide a principled rather than an *ad hoc* response to questions about the number of actions that were performed. Since the four descriptions are logically independent of one another, it follows from (Q) that we do not have one action described in four

ways but four actions performed at the same time by the same person.

The question whether there were four actions or only one would be utterly trivial if it were not for the fact that an important issue is at stake. On the assumption that there is only one action described in four different ways, Davidson asserts that statements which provide reasons for action are quasi-intentional; by this he means, apparently, that the law of the substitution of identicals fails for such statements. His argument is as follows: suppose I flip on the switch with the intention of turning on the light, but, unknown to me, there is a prowler in the room whom I frighten. Thus it is true that:

- (1) My reason for flipping the switch was that I wanted to turn on the light.

As we have seen Davidson also assumes that:

- (2) Flipping the switch is the same action as alerting the prowler.

By the law of substitution of identicals we can infer:

- (3) My reason for alerting the prowler was that I wanted to turn on the light.

Since (3) is false, while (1) and, so it is alleged, (2) are true, what is presumably at fault here is the law of substitution.

On Davidson's analysis the deduction of (3) from (1) and (2) is a case of invalid reasoning since a false conclusion has been inferred from true premisses. But on the view defended in this paper, the reasoning is not invalid at all; what is at fault is simply that (2) is false so that we do not have a case of deriving a false conclusion from true premisses. In recent years, it has become common to raise doubts about the unrestricted validity of the law of substitution of identicals.⁹ In this discussion, I shall not try to resolve these doubts in general. But I need to point out that the doubt raised about the law with respect to action statements is based upon an incorrect view of the metaphysics of action. The fact that the alternative metaphysics I have here been defending preserves the law to this extent is another argument in its favor.

Hunter College, The City University of New York

Received May 17, 1968

⁸ For a criticism of principle (Q) similar to the one rebutted here see Richard Brandt and Jaegwon Kim, "The Logic of the Identity Theory," *The Journal of Philosophy*, vol. 64 (1967), p. 526.

⁹ A classic statement is ch. 8 of W. V. Quine's *From a Logical Point of View* (New York, 1961). In an unpublished paper, Professor Howard Kahane argues cogently that a more careful statement of the inferences which the law licenses does not lead to doubts of its validity.

X. DEFINING ART

GEORGE DICKIE

IN recent years it has been argued that the expression "work of art" cannot be defined and Morris Weitz has even argued that *being an artifact* is not a necessary condition for being a work of art.¹ More recently, however, Joseph Margolis has offered a definition² and Maurice Mandelbaum has made tentative suggestions about defining "art."³

I shall not repeat the well-known argument of Weitz, whose views I take to be representative of those who maintain that "art" cannot be defined, but shall state his main conclusion and comment on one of his arguments. Neither shall I repeat the arguments of Margolis or Mandelbaum, but I do want to note (1) that they agree that artifactuality is a necessary condition of art, and (2) that Mandelbaum points out the significance of the *non-exhibited* characteristics of art for the definition of "art."

Weitz's main conclusion is that there are no necessary and sufficient conditions for the definition of "art" or for any of the subconcepts of art, such as "novel," "tragedy," "painting," and so on. All of these notions are open concepts and their instances have "family resemblances."

Weitz rejects artifactuality as a necessary condition of art because we sometimes make statements such as "This driftwood is a lovely piece of sculpture."⁴ We do sometimes speak this way of natural objects, but nothing follows from this fact. Weitz is confused because he takes the driftwood remark to be a descriptive statement and it is not. Weitz himself, quite correctly, distinguishes between an evaluative use and a descriptive use of "work of art,"⁵ and once this distinction is understood it can be seen that the driftwood remark is an evaluation of the driftwood. But it is, of course, the descriptive sense of "work of art" which is at issue

when the question of whether "art" can be defined is raised. I maintain that the descriptive use of "work of art" is used to indicate that a thing belongs to a certain category of artifacts. By the way, the evaluative sense can be applied to artifacts as well as nonartifacts, as when we say, "That painting is a work of art." Such remarks are not intended as tautologies.

Before going on to discuss the second condition of the definition of the descriptive sense of "art," it will be helpful to distinguish the generic concept of art from the various subconcepts which fall under it. It may very well be the case that all or some of the subconcepts of art, such as novel, tragedy, ceramics, sculpture, painting, and so on, may lack necessary and sufficient conditions for their application as subconcepts and it still be the case that "work of art," which is the genus of all these subconcepts, can be defined. For example, there may not be any characteristics which all tragedies have which would distinguish them from comedies, satyr plays, happenings, and the like within the domain of art. Even if this were the case, in the light of the foregoing, tragedies and all other works of art would have at least one characteristic in common, namely, artifactuality. Perhaps artifactuality and some one or more other features of works of art distinguish them from nonart. If all or some of the subconcepts of art cannot be defined and, as I think is the case, "art" can be, then Weitz is right in part.

* * *

Assuming that artifactuality is the genus of art, the differentia is still lacking. This second condition will be a social property of art. Furthermore, this

¹ Morris Weitz, "The Role of Theory in Aesthetics," *The Journal of Aesthetics and Art Criticism*, vol. 15 (1956), pp. 27-35; reprinted in *Philosophy Looks at the Arts*, ed. by Joseph Margolis (New York, 1962); Paul Ziff, "The Task of Defining a Work of Art," reprinted in *Aesthetics and the Philosophy of Criticism*, ed. by Marvin Levich (New York, 1963); William Kennick, "Does Traditional Aesthetics Rest on a Mistake," *Mind*, vol. 66 (1958), pp. 317-334.

² *The Language of Art and Art Criticism* (Detroit, 1965), pp. 37-47. Margolis' definition is not satisfactory, however; see Andrew Harrison's review in *Philosophical Books*, vol. 7 (1966), p. 19.

³ "Family Resemblances and Generalization Concerning the Arts," *American Philosophical Quarterly*, vol. 2 (1965), pp. 219-228.

⁴ *Op. cit.*, p. 57.

⁵ *Ibid.*, p. 56.

social propriety will, in Mandelbaum's terminology, be a nonexhibited, relational property.

W. E. Kennick contends that such an approach to the definition of "art" is futile. He argues from such facts as that the ancient Egyptians sealed up paintings and sculptures in tombs to the conclusion that "The attempt to define Art in terms of what we do with certain objects is as doomed as any other."⁶ There are several difficulties with Kennick's argument. First, the fact that the Egyptians sealed up paintings and sculptures in tombs does not entail that they generally regarded them differently from the way in which we regard them. Indeed, they might have put them there for the dead to appreciate, or simply because they belonged to the dead person, or for some other reason. The Egyptian practice does not prove a radical difference between their conception of art and ours such that a definition which subsumes both is impossible. Secondly, there is no need to assume that we and the ancient Egyptians (or any other group) share a common conception of art. I would be happy to be able to specify the necessary and sufficient conditions for the concept of art which we have (we present-day Americans, we present-day Westerners, we Westerners since the organization of the system of the arts in or about the 18th century—I am not sure of the exact limits of the "we"). Kennick notwithstanding, we are most likely to discover the differentia of art by considering "what we do with certain objects," that is, "works of art." But, of course, there is no guarantee that any given thing we or an ancient Egyptian might possibly do with a work of art will throw light on the concept of art. Not every *doing* will reveal what is required.

Arthur Danto's stimulating article, "The Artworld,"⁷ is helpful here. In speaking of Warhol's Brillo Carton and Rauschenberg's Bed, he writes, "To see something as art requires something the eye cannot de[s]cry—an atmosphere of artistic theory, a knowledge of history of art: an artworld."⁸ What the eye cannot descry is a complicated non-exhibited characteristic of the artifacts in question. The "atmosphere" of which Danto speaks is elusive, but it has a substantial content. Perhaps this content can be captured in a definition. I shall first state the definition and then go on to defend it.

A work of art in the descriptive sense is (1) an artifact (2) upon which some society or some sub-group of a society has conferred the status of candidate for appreciation.

The definition speaks of the conferring of the status of *candidate* for appreciation: nothing is said about actual appreciation and this leaves open the possibility of works of art which, for whatever reason, are not appreciated. Also, not every aspect of a work is included in the candidacy for appreciation, for example, the color of the back of a painting is not ordinarily an object of appreciation. The problem of *which* aspects of a work of art are to be included within the candidacy for appreciation is a question which I have pursued elsewhere.⁹

Just how is the status of candidate for appreciation conferred? An artifact's hanging in an art museum, a performance at a theater, and the like are sure signs that the status *has been conferred*. But many works of art never reach museum walls and some are never seen by anyone but the artist himself. The status, therefore, must be conferrable by a single person's treating an artifact as a candidate for appreciation, usually the artist himself, although not always, because someone might create an artifact without ever considering it as a candidate for appreciation and the status be conferred by some other person or persons. But can status be conferred so easily? We associate status with ceremony—the wedding ceremony and the status of being married, for example. However, ceremony is not the only way of getting married, in some jurisdictions common-law marriage is possible—a status acquired without ceremony. What I want to suggest is that, just as two persons can acquire the status of common-law marriage within a legal system, an artifact can acquire the status of a candidate for appreciation within the system which Danto has called "the artworld."

A number of questions arise about this notion of status of candidate for appreciation and perhaps the whole matter can best be clarified by stating them and trying to answer them. Probably the first question is: what *kind* of appreciation? Surely the definition does seem to suggest that there is a special kind of "aesthetic" appreciation. Appreciation is not crucial, but something should

⁶ Kennick, *op. cit.*, p. 330.

⁷ *The Journal of Philosophy*, vol. 61 (1964), pp. 571-584.

⁸ *Ibid.*, p. 580.

⁹ In my "Art Narrowly and Broadly Speaking," *American Philosophical Quarterly*, vol. 5 (1968), pp. 71-77, where I analyze the notion of *aesthetic object*. The subject of the present essay is the concept of *art* which, although related to the notion of *aesthetic object*, is distinct from it.

be said about it to prepare the way for the crucial point. The kind of appreciation I have in mind is simply the kind characteristic of our experiences of paintings, poetry, novels, and the like. This remark seems to collapse the definition into circularity, but it does not because "work of art" (the term defined) does not appear in the explanation of appreciation, only subconcept terms appear. Another apparent problem is that works of art differ so much from one another—for example, comedies are very different from tragedies—that it seems unlikely that the appreciation characteristic of our experience of one kind of work has something in common with the appreciation characteristic of our experience of another kind of work. But paintings, poems, and plays are the *objects* of our appreciation and the fact that the objects differ considerably does not mean that the various appreciations differ. Indeed, if we mean by "appreciation" something like "in experiencing the qualities of a thing one finds them worthy or valuable," then there is no problem about the similarity of the various appreciations.

It can now be seen that appreciation will not serve to pick out the subclass of works of art from the class of artifacts—it is too broad: many artifacts which are obviously not works of art are appreciated. To pick out the class of works of art one must stress the conferring of the status of candidate rather than appreciation. When, for example, a salesman of plumbing supplies spreads his wares before us, he presents them for our appreciation all right, but the presenting is not a conferring of status of candidate, it is simply a placing before us. But what is the difference between "placing before" and "conferring the status of candidate?" The difference is analogous to the difference between my uttering "I declare this man to be a candidate for alderman" and the head of the election board uttering the same sentence while acting in his official capacity. When I utter the sentence it has no effect because I have not been vested with any authority in this regard. Of course the analogy is not a complete one—lines of authority in the politico-legal world are by and large explicitly defined and incorporated into law, while lines of authority (or something like authority) in the artworld are nowhere codified. The artworld carries on its business at the level of customary practice. Still there is a practice and this defines a social institution. To return to the plumbing line, the salesman's presentation is different from Duchamp's superficially similar act of placing a urinal which he

christened "Fountain" in that now famous art show. The point is that Duchamp's act took place within a certain institutional setting and that makes all the difference. Our salesman of plumbing supplies could do what Duchamp did, that is, convert a urinal into a work of art, but he probably would not—such weird ideas seem to occur only to artists with bizarre senses of humor. Please remember that when I say "Fountain" is a work of art, I am not saying it is a good one. And in making this last remark I am not insinuating that it is a bad one either.

Duchamp's "ready-mades" raise the question—"If urinals, snowshovels, and hatracks can become works of art, why can't natural objects such as driftwood become works of art?" and, of course, driftwood and other natural objects can become works of art if any one of a number of things is done to them. One thing which would do the trick would be to pick it up, take it home, and hang it on the wall. Another thing which would do the trick would be to pick it up and enter it in an exhibition. (I was, by the way, assuming that Weitz's sentence about driftwood referred to a piece of driftwood in its ordinary situation on a beach and untouched by human hand.) This means that natural objects which become works of art acquire their artifactuality (are artifactualized) at the same time that the status of candidate for appreciation is conferred on them. But perhaps a similar thing ordinarily happens with paintings, poems, and such; they come to exist as artifacts at the same time that they have conferred on them the status of candidate for appreciation. (Of course, being an artifact and being a candidate for appreciation are not the same thing—they are two properties of a single thing which may be acquired at the same time.) A somewhat more complicated case would be an artifact from a primitive culture which played a role in a religious system and which had no artistic function in the sense developed here. Such an artifact might become a work of art in our culture in a way similar to that in which driftwood might become a work of art. However, such a religious object which becomes a work of art would be an artifact in two senses, but the driftwood in only one. (I am not suggesting that something cannot be a religious object and work of art at the same time—there are many counter-instances to this in our own culture.)

A question which frequently arises in connection with discussions of the concept of art is "How are we to conceive of paintings done by individuals such

as Betsy the chimpanzee from the Baltimore Zoo?" It all depends on what is done with the paintings. (Note that I unhesitatingly call the objects paintings, although I am uncertain about their status as works of art.) For example, The Field Natural History Museum in Chicago recently exhibited some chimpanzee paintings. In the case of these paintings we must say that they are not works of art. However, if they had been exhibited a few miles away at the Chicago Art Institute they would have been works of art. (If, so to speak, the director of the Art Institute had gone out on a limb.) It all depends on the institutional setting.

In concluding, it may be worthwhile to consider in what ways the definition offered here differs from some traditional definitions. (1) It does not attempt to smuggle a conception of good art into the definition of "art." (2) It is not, to use Margolis' term, "overloaded," as is the one Margolis cites as a horrible example: "Art is a human activity which explores, and hereby creates, new reality in a

suprarational, visional manner and presents it symbolically or metaphonically,¹⁰ as a micro-cosmic whole signifying a macrocosmic whole."¹¹ (3) It does not contain any commitment to any metaphysical or unempirical theory, as contrasted with, for example, the view that art is unreal. (4) It is broad enough so that those things generally recognized as art can be brought under it without undue strain, as contrasted with, for example, the imitation definition which involves enormous strain in trying to show that every work of art is an imitation of something or other. (5) It takes into account (or at least attempts to) the actual practices of the artworld of the past and of the present day.

Now what I have been saying may sound like saying, "a work of art is an object of which someone has said, 'I christen this object a work of art'." And I think it is rather like that. So one *can* make a work of art out of a sow's ear, but of course that does not mean that it is a silk purse.¹²

University of Illinois at Chicago Circle

Received March 18, 1968

¹⁰ There are apparently two typographical errors here. Margolis quotes the word as "metaphonically" and the original text reads "metaphonically." A reading of the original text indicates that it should have been "metaphorically."

¹¹ *Op. cit.*, p. 44. The passage is quoted from Erick Kahler's "What is Art?" in *Problems in Aesthetics*, ed. by Morris Weitz (New York, 1959).

¹² Thanks are due to Monroe Beardsley, Marcia Eaton, William Hayes, Arnold Levison, and Maurice Mandelbaum who read this paper in manuscript and made many helpful suggestions.

PRINTED IN ENGLAND
by C. Tinling & Co. Ltd., Liverpool, London and Prescott

AMERICAN PHILOSOPHICAL QUARTERLY

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this is to include occasional collections of original articles on a common theme. These monographs can be purchased separately but are available to individual subscribers at *half price* (though not to institutional subscribers).

No. 1. STUDIES IN MORAL PHILOSOPHY. *Contents:* Kai Nielsen, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supererogation and Duties"; Lawrence Haworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, \$6.00.

No. 2. STUDIES IN LOGICAL THEORY. *Contents:* Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Honderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, \$6.00.

No. 3. STUDIES IN THE PHILOSOPHY OF SCIENCE. *Contents:* Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What Are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, \$6.00.

AMERICAN PHILOSOPHICAL QUARTERLY

Edited by

NICHOLAS RESCHER

With the advice and assistance of the Board of Editorial Consultants:

Alan R. Anderson

Kurt Baier

Stephen F. Barker

Monroe Beardsley

Nuel D. Belnap, Jr.

Roderick M. Chisholm

L. Jonathan Cohen

James Collins

James M. Edie

José Ferrater-Mora

Peter Thomas Geach

Adolf Grünbaum

Carl G. Hempel

John Hospers

Raymond Klibansky

Hugues Leblanc

Ernan McMullin

Benson Mates

John A. Passmore

Günther Patzig

Richard H. Popkin

Richard Rorty

George A. Schrader

Michael Scriven

Wilfrid Sellars

Alexander Sesonske

Manley H. Thompson, Jr.

James F. Thomson

John W. Yolton

Volume 6/Number 4

OCTOBER 1969

CONTENTS

- | | | | |
|---|-----|--|-----|
| I. HECTOR-NERI CASTANEDA: <i>Ought, Value, and Utilitarianism</i> | 257 | VII. ROLF SARTORIUS: <i>A Neglected Aspect of the Relationship Between Berkeley's Theory of Vision and His Immaterialism</i> | 318 |
| II. MICHAEL STOCKER: <i>Consequentialism and Its Complexities</i> | 276 | VIII. ANTONIO S. CUA: <i>Morality and the Paradigmatic Individuals</i> | 324 |
| III. ARTHUR N. PRIOR: <i>Tensed Propositions as Predicates</i> | 290 | IX. GUIDO KÜNG: <i>The Role of Language in Phenomenological Analysis</i> | 330 |
| IV. THOMAS MAUTNER: <i>Aquinas's Third Way</i> | 298 | X. J. E. BARNHART: <i>Human Rights as Absolute Claims and Reasonable Expectations</i> | 335 |
| V. W. GREGORY LYCAN: <i>On Intentionality and the Psychological</i> | 305 | XI. <i>Corrigenda</i> | 341 |
| VI. JOHN E. ATWELL: <i>Husserl on Signification and Object</i> | 312 | | |

AMERICAN PHILOSOPHICAL QUARTERLY

POLICY

The *American Philosophical Quarterly* welcomes articles by philosophers of any country on any aspect of philosophy, substantive or historical. However, only self-sufficient articles will be published, and not news items, book reviews, critical notices, or "discussion notes."

MANUSCRIPTS

Contributions may be as short as 2,000 words or as long as 25,000. All manuscripts should be typewritten with wide margins, and at least double spacing between the lines. Footnotes should be used sparingly and should be numbered consecutively. They should also be typed with wide margins and double spacing. The original copy, not a carbon, should be submitted; authors should always retain at least one copy of their articles.

COMMUNICATIONS

Articles for publication, and all other editorial communications and enquiries, should be addressed to: The Editor, *American Philosophical Quarterly*, Department of Philosophy, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

OFFPRINTS

Authors will receive gratis two copies of the issue containing their contribution. Offprints can be purchased through arrangements made when checking proof. They will be charged for as follows: The first 50 offprints of 4 pages (or fraction thereof) cost \$12, increasing by \$1 for each additional 4 pages. Additional groups of 50 offprints of 4 pages cost \$8, increasing by \$1 for each additional 4 pages. Covers will be provided for offprints at a cost of \$4 per group of 50.

SUBSCRIPTIONS

The price *per annum* is six dollars for individual subscribers and ten dollars for institutions. Checks and money orders should be made payable to the *American Philosophical Quarterly*. All back issues are available, and are sold at the rate of two dollars to individuals, and three dollars to institutions. All correspondence regarding subscriptions and back orders should be addressed directly to the publisher (Basil Blackwell, Broad Street, Oxford, England).

MONOGRAPH SERIES

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this includes occasional collections of original articles on a common theme. These monographs can be purchased separately but are available to individual subscribers at a substantially reduced price. The back cover of the journal may be consulted for details.



I. OUGHT, VALUE, AND UTILITARIANISM

HECTOR-NERI CASTANEDA

MANY different kinds of utilitarianism have been offered in the philosophical market. What they all have in common is their being theories about the connection between value and deontic modalities (rightness, wrongness, and oughtness). There is, apparently, such a connection. Yet it is wholly unclear what the connection is, as the very multiplicity of utilitarianisms bears witness. It is clear, however, that any theory about the connection between value and deontic modalities must satisfy the crucial test of conforming to all the entailments or implication relationships between deontic statements.

Our immediate purpose here is to subject act-utilitarianism to this deontic test. Rule-utilitarianism will be discussed in another installment of this study. Our ultimate purpose is to shed some light on the connection between value and deontic modalities.

In spite of the traditional criticisms of act-utilitarianism, it is still strongly supported by many philosophers and many non-philosophers alike. Many of us strongly incline to agree with W. D. Ross, [1], pp. 21f,¹ that it is a source of genuine *prima facie* obligations, in the midst of other sources. It is a beautiful and widely favored view on the nature of the solution to conflicts of duties. Thus, the study of the logic of act-utilitarianism is also important for its own sake.

The main results of the investigation are: (i) no form of simple act-utilitarianism passes the deontic test; (ii) more strongly: merely maximizing value is in general neither a necessary nor a sufficient condition even for the *prima facie* rightness, or the *prima facie* obligatoriness, of an action. To show this 49 utilitarian theses are examined. Among the important by-products are the development of two concepts of alternative action and the discussion of the method for measuring the value of disjunctive acts. After a discussion of probabilistic utilitarianism, the paper concludes with a formulation of a *complex* utilitarianism that both combines pure value considerations with probabilities and seems deontically viable.

I. PRELIMINARY TEST OF UTILITARIANISM

A characteristic act-utilitarian principle scheme is this:

(O') X ought _{i} to do A in circumstances C , if and only if X 's doing A in C will bring about a greater balance of good of kind i over bad of kind i than X 's doing any other action open to him in C . (The specific import of the subscript i will be explained in Sect. II.)

(O') covers a very wide utilitarian territory. The kind i of good can be intrinsic good, or extrinsic, or instrumental, or consequential good, or a total balance of all these types of good. It can be the agent's happiness, or a group's happiness, or the happiness of the greater number of people, or the most satisfactory distribution of happiness among all human beings with respect to some chosen principles, e.g., of justice of rewards and punishments. The oughtness at issue in (O') may be some specific or generic moral, or legal or political, oughtness. The connective "if and only if" may express a definitional or an otherwise necessary biconditional, or a material biconditional.

A general principle of deontic logic widely accepted is this:

(cO'₃) " X ought morally (legally, -ly) to do A & B " entails
" X ought morally (legally, -ly) to do A " & " X ought morally (legally, -ly) to do B ."

In the statement of (cO'₃) we have used '&' instead of "and" to indicate most emphatically that we are dealing with logical conjunction, and not with other relations expressed by the word "and." It would be absurd to claim that (cO'₃) holds for other logical operations "and" may normally express.

Suppose that a man hereafter called *Utilius* is in circumstances C and ought _{i} (in the relevant sense or type of oughtness) to do a conjunctive act A & B . By cO'₃ *Utilius* ought _{i} to do A , and *Utilius* ought _{i} to do B . By (O'), *Utilius* doing A .

¹ Numeral within square brackets refers to the items in the bibliography at the end of the study.

will bring about a greater balance of good (happiness, pleasure, etc.) than his doing any other action open to him in his circumstances C , *a fortiori*, than his doing B . Likewise, by (O') Utilius' doing B will bring about a greater balance of good (happiness, pleasure, etc.) than his doing any other action he can do in C , *a fortiori*, than his doing A . Thus, Utilius' doing A will bring about *both a greater and a lesser* balance of good (happiness, pleasure, etc.) than his doing B ! This is a contradiction in terms. Hence, (O') is false, regardless of whether it is put forward as an analysis of the meaning of "ought" or as an analysis of moral, legal, or political obligation, or as a synthetic principle of action, and regardless of whether the act-utilitarianism involved is hedonistic, eudaimonistic, or agathistic.

A simple inspection of the preceding reasoning reveals that what leads to the self-contradiction is the only-if half of (O'). Thus we seem to have here a proof that utilitarian reasons cannot be necessary conditions for obligatoriness.

But before we start counting the results hatched off by the preceding argument we must consider an immediate and very natural reply. It may be replied that (O') is really a *naïve* type of utilitarianism and not what most utilitarians have meant, even if they have sometimes expressed themselves in ways that sound like (O'). What they mean to compare is *not* an action A with all other actions open to the agent, but an action A with all the other alternative actions open to the agent in his circumstances. *It is part and parcel of action B being an alternative to action A*, the reply continues, *that action A and action B are incompatible*. Thus, the reply proceeds, if Utilius ought to do $A \& B$, then his doing A is not incompatible with his doing B , so that A and B are not alternatives. Therefore, you are misapplying (O') when you compare the balance of good (happiness, etc.) brought about by Utilius' doing A with the balance brought about by his doing B . What you should attack is not (O') but:

- (O'') X ought _{i} to do A in circumstances C , if and only if X 's doing A in C will bring about a greater balance of good of kind i over bad of kind i , than X 's doing any other alternative action open to X in circumstances C .

And, the reply concludes, as long as you have not touched (O'') utilitarianism is unaffected.

This is a very illuminating reply. It demands a

deeper look into the logic of utilitarianism. To grasp the problem better consider a preliminary rejoinder as follows:

- (1) Let Utilius ought _{i} to do $A \& B$.

By (O''), Utilius' doing $A \& B$ would bring about a greater balance of value (of whatever kind is at issue) than his doing any incompatible action he can do in his circumstances. Hence, in obvious abbreviational form:

- (2) $\text{Value}(A \& B) > \text{Value}(A \& \sim B),$
 $\text{Value}(\sim A \& B), \text{Value}(\sim A \& \sim B).$

Now another principle governing ought is roughly this: "If action A entails action B , then X ought _{i} to do A entails X ought _{i} to do B ." Hence, from (1) it follows that:

- (3) Utilius ought _{i} to do $(A \& B \vee A \& \sim B).$

The incompatible alternatives to this disjunction are $\sim A \& B$ and $\sim A \& \sim B$. Hence by (O''):

- (4) $\text{Value}(A \& B \vee A \& \sim B) > \text{Value}(\sim A \& \sim B).$

The following principle of value theory has been endorsed and defended by von Wright in [2], pp. 26, 37, and elsewhere, where p , q , and r are conjunctions of simpler actions or their negations, but not both:

- (VW) If $\text{Value}(p \vee q) > \text{Value}(r)$, then both
 $\text{Value}(p) > \text{Value}(r)$ and $\text{Value}(q) > \text{Value}(r).$

Hence, from (4) and (VW):

- (5) $\text{Value}(A \& \sim B) > \text{Value}(\sim A \& B).$

By paralleling the steps from (3) to (5), given that action $A \& B$ entails action $A \& B \vee \sim A \& B$, we derive:

- (8) $\text{Value}(\sim A \& B) > \text{Value}(A \& \sim B).$

Clearly, (5) and (8) exhibit the same kind of contradiction encountered in our preliminary argument. Indeed, (cO'₃) is also in the background, since action A is equivalent to $A \& B \vee A \& \sim B$ and B is equivalent to $A \& B \vee \sim A \& B$.

The weakest link in the preceding rejoinder is the move from step (4) to step (5). The utilitarian may certainly counter-reply by rejecting von Wright's principle (VW), even though the application of (VW) made above is particularly plausible: the disjuncts are mutually incompatible and each disjunct is a complete action that Utilius can perform in the circumstances at issue. Nevertheless, we are prepared to side with the utilitarian against von Wright. Von Wright's principle leads to the incomparability of $A \vee B$ with C , where, say, $\text{Value}(A) = 1,000$; $\text{Value}(B) = 20$, and $\text{Value}(C) = 20$. In a case like this it seems to be more

satisfactory to say that the value of $A \vee B$ is greater than the value of C .

We submit, therefore, that our preliminary rejoinder does not rebut the reply to our preliminary argument. The reply does not, however, protect all types of utilitarianism from the preliminary argument. (See footnote 5 below.) But it protects the most interesting and sophisticated ones. The ones that fall down are definitional utilitarianisms. Once we introduce the requirement that the variable ' A ' in (O) ranges over actions that belong to a set of alternatives, (O) fails to provide an analysis of the unrestricted meaning of "ought" as we use it, where " X ought to do A " has a meaning even if action A is not taken from a set of alternatives in the sense required by the utilitarian philosopher.

Definitional act-utilitarianisms are, therefore, ruled out. Yet there is a large troupe of synthetic utilitarianisms, *a priori* or otherwise, awaiting to be tested. But before we test them we must clarify the notion of open alternative action, and we must discuss the assignment of value to disjunctions. We must also be clearer on the principles of deontic logic we are assuming. We shall clarify all these things in the next four sections of this paper.

II. DEONTIC LOGIC

There are many different types or species of ought or of obligation. There are even many different senses, all relevant to practical thinking, in which acts are said to be obligatory. Not only do we contrast legal with moral obligations, for instance, but we also face conflicts of obligations within the realm of the law, or of morality, itself. The logical analysis of a mere conflict of promises (or contracts) shows that we must reckon with an indefinitely large multiplicity of oughts (or obligations or duties): one ought for each of the conflicting promises, and one *overriding ought* that solves the conflict. The proper analysis of a statement like "Smith, inasmuch as you promised Jones to wait for his friend, you ought to wait for Jones's friend" is "Smith, you made a promise, call it j , to Jones that you would wait for his friend, and you ought _{j} to wait for him." Here the subscript ' j ' signals the species of ought involved, in the same way in which often the subscript ' i ' signals the species of blueness predicated with the adjective "blue _{i} ." (For the

analysis of conflicts of duties see [4], [9], and [11].) From now on we shall use "ought _{i} " to refer to each type of ought, including the overriding ought, which will be referred to with "ought _{i} ."

The oughts introduced by the several types of utilitarianism are, clearly, not identical with the overriding ought. But we shall not discuss their relationship. Here we cannot detain ourselves to discuss the criteria for determining types or species of oughts, or practical senses of "ought." We must note, however, that regardless of how one specifies the types of ought or individuates the practical senses of "ought," all practical oughts have a characteristic formal structure in common, which is a necessary condition of their practicality. Palpably, that common structure is more fundamental to practical concepts and words than the idiomatic or dialectical differences between the words "ought," "obligation," "duty," and "requirement." For our present purpose the difference in the uses of these words may be taken to come in the subscript ' i ' attached to "ought" or to "obligatory."

Here we simply cannot defend in detail each of the basic principles of deontic logic which we shall be using in the ensuing examination of utilitarianism. This defense has already been carried out elsewhere, especially in [4], [6]-[8], [11], which are, thus, an integral part of this paper; they go together with [3], [5], and [13].

The main principles of deontic logic which we shall need here are:

- (O₂) " X ought _{i} to do A " entails "It is right _{i} for X to do A ."
- (O₃) " X ought _{i} to do A & X ought _{i} to do B " entails " X ought _{i} to do A & B ."
- (O₄)² " p & X ought _{i} to do A " entails " X ought _{i} to do p & A ."
- (DR₁) If command " X , do A " entails " X , do B ," then " X ought _{i} to do A " entails " X ought _{i} to do B ."
- (DD₁) "It is right _{i} for X to do A " entails, and is entailed by, "It is not the case that X ought _{i} to do not- A ."
- (DD₂) "It is wrong _{i} for X to do A " entails, and is entailed by, "It is not right _{i} for X to do A ."

Naturally, we adopt a generalized command form of *modus ponens*:

- (MP) From α and "if α , β " infer β .

² This principle is at first sight somewhat counter-intuitive. For this reason it is used only once in this essay, in connection with probabilistic utilitarianism in Sect. VIII.

These are the simplest cases of axioms and rules constituting calculus (O^*), formulated in [4] and [11]. In (O_3) and (O_4) '&' stands for "and" when it is used as a mere logical conjunction. Obviously, neither (O_3) nor (O_4) hold in general for other uses of "and." (On this matter see [4], [5], [6] and [13].) In (O_4) ' p ' is a propositional variable, and it ranges over statements, not actions. An example of (O_4) is this: "Peter has a sister and he ought_i to bring her in" entails "Peter ought_i to bring his sister in." Another example is: "Jones hates Smith, but Smith ought_i to be nice to him" entails "Smith ought_i to do the following: be nice to Jones, who hates him." A third illustration is furnished by the necessarily true conditionals of the form "If it is the case that p and doing A is obligatory, then doing A while it is the case that p is obligatory."

(DR_1) is a trouble-free formulation of the principle ordinarily put as follows (aside from the subscripts, naturally):

(DR_1) If doing A entails doing B , then " A ought_i to be done" entails " B ought_i to be done."

But (DR_1) raises problems of interpretation. The gerundial locution "doing A " is not clear, and it allows of its being interpreted as a statement about the performance of A . Thus, it allows that (DR) below be a genuine instance of (DR_1):

(DR_1) If " X performs A " entails " X performs B ," then " X ought_i to do A " entails " X ought_i to do B ."

(DR_1) leads immediately to the so-called paradox of the Good Samaritan (as shown in [4], pp. 33-34 and [7], pp. 13-15). For reasons discussed in these papers, "doing A " and "doing B " in (DR_1) must be interpreted as expressing prescriptions, or commands, orders, requests, petitions, or pieces of advice. Similarly, this interpretation is required in order to avoid the problem that Chisholm raised in [16] for one-sorted deontic calculi.³ (See [7] for this and other related points.)

By means of (MP) and (DR_1) (and the principles of the logic of commands or prescriptions, for which see [4] and [7]), we can derive the converses of (O_3) and (O_4). Of course, we can regard the converses themselves as axiomatic:

(cO_3) " X ought_i to do A & B " entails " X

ought_i to do A " and " X ought_i to do B ."
(cO_4) " X ought_i to do p & A " entails both that p and " X ought_i to do A ."

From (O_2) and (DD_1), by means of (MP) and (DR_1), we can also derive:

(O_5) " X ought_i to do A " entails that " X , do A " is self-consistent.

III. ALTERNATIVE ACTIONS

Let us call *utilitarian actions* the alternative actions in whose comparisons of value (whether intrinsic, or extrinsic, or instrumental, or total) utilitarianism is interested. As we have seen, alternative utilitarian actions are incompatible with each other. This suggests that a set of utilitarian alternatives is a set of conjunctions of certain primary actions or their negations, but not both. Suppose that in a given situation Utilius is considering what to do, and that the primary, non-analyzed actions that he is concerned with are A_1, \dots, A_{n-1}, A_n . We assume, not that these primary acts are intrinsically unanalyzable, but only that given Utilius' *circumstances* they need not be analyzed. Then, using braces, as is customary, to name a set in term of its members, according to the present suggestion, the maximal set of utilitarian alternatives for Utilius in that situation is:

$$\begin{aligned} \alpha_1(A_1 \dots A_n) = \{ & A_1 \& \dots \& A_{n-1} \& A_n, \\ & A_1 \& \dots \& A_{n-1} \& \sim A_n, A_1 \& \dots \& \sim A_{n-1} \& A_n, \\ & A_1 \& \dots \& \sim A_{n-1} \& \sim A_n, \dots, \\ & \sim A_1 \& \dots \& \sim A_{n-1} \& \sim A_n \}. \end{aligned}$$

More likely than not, Utilius will not be able to perform each one of the alternatives in $\alpha_1(A_1 \dots A_n)$. The set of utilitarian alternatives actually open to him would then be a subset $\alpha_n(A_1 \dots A_n)$ of $\alpha_1(A_1 \dots A_n)$. But we shall postpone the discussion of openness to Sect. IV. In any case, his open alternatives are both incompatible by pairs and jointly exhaustive of all the possibilities of action as far as Utilius is concerned in that situation. Thus, by comparing the amounts of value that would accrue to the universe if he were to perform each of those alternatives, Utilius would, it seems, be comparing all that he need compare to find out what he ought_i to do.

³ For other treatments of the Good Samaritan paradox and the Chisholm puzzle for deontic calculi see Sellars [23] and Aqvist [24]. My solution in [4] and [7] is still the simplest one. As Chisholm himself noted, his puzzle does not arise in the calculus formulated in [8]. The criticisms against [7] and [11] that Lawrence Powers raised in [25] pp. 390-395, were anticipated in [4]. They apply only to certain examples given in [7] for the distinction between action prescriptively considered and action considered as a circumstance, but not to the distinction itself or the final examples in [4] or [7].

The above suggestion is undoubtedly meritorious. Yet we cannot, without more ado, assume that (O'') makes obligatory_i the member of a set $\alpha_h(A_1 \dots A_n)$ which has the highest value. In some situations we ought_i to do $(A_1 \& A_2 \vee A_1 \& \sim A_2)$, even though it is not the case that we ought_i to do one disjunct rather than the other. Thus, it seems, we must, if we are utilitarians, compare the value of $A_1 \& A_2 \vee A_1 \& \sim A_2$ with the values of $A_1 \& A_2$ and $A_1 \& \sim A_2$, even though the former is incompatible with neither of the latter. Yet such a comparison must not be allowed; otherwise the reply to our preliminary argument in Sect. I collapses and the argument is reinstated in its full force. We must, then, let (O'') refer to sets other than subsets of $\alpha_1(A_1 \dots A_n)$.

There are two ways of accommodating disjunctive obligations. The first way is faithful to the original idea of a set of utilitarian alternatives being a set of actions incompatible by pairs and jointly exhaustive of the possibilities of action open to the agent in the given circumstances. It is as follows. Take a set $\alpha_1(A_1 \dots A_n)$. Its members will be called *basic utilitarian actions* or *buas*. Disjunctions of *buas* will be called *utilitarian disjunctions*. *Buas* and their disjunctions will all be called *utilitarian actions*. Now, a set $\alpha_k(A_1 \dots A_n)$ of utilitarian actions all of whose members are incompatible by pairs and jointly exhaustive of the open possibilities of action will be called a set of α - or α_k -alternatives; if A and B are members of one set $\alpha_k(A_1 \dots A_n)$ they will be said to be α - or α_k -alternatives. We shall take the relation of α -alternatives to be reflexive, symmetric, and transitive. When there is no danger of confusion we shall drop the references to the unanalyzed actions from which we obtain our utilitarian actions, writing simply ' α_1 ' or ' μ_h ', or ' μ ' or ' μ^k ' (to be introduced below).

To illustrate α -alternatives, consider two primary actions A and B . They determine as set of *buas* $\alpha_1(AB)$, and as possible sets of open α -alternatives the following:

$$\begin{aligned}\alpha_1(AB) &= \{ [A \& B, A \& \sim B, \sim A \& B, \\ &\quad \sim A \& \sim B] \} \\ \alpha_2(AB) &= \{ [A \& B, A \& \sim B, \sim A \& \sim B] \} \\ \alpha_3(AB) &= \{ [A \& B \vee \sim A \& \sim B, \sim A \& B] \} \\ \alpha_4(AB) &= \{ [A \& B \vee \sim A \& \sim B, A \& \sim B \vee \\ &\quad \sim A \& \sim B] \}, \text{ etc.}\end{aligned}$$

Clearly, if we require that (O'') apply to each set of α -alternatives separately, we may very well find that the disjunction, say, $A \& B \vee \sim A \& \sim B$ is obligatory_i, without having to compare its

value with the values of actions not incompatible with it. This would be the case if, for instance, that disjunction exceeds in value the other member of $\alpha_3(AB)$ or $\alpha_4(AB)$. Thus, in order not to disturb the reply to our preliminary argument in Sect. I we interpret (O'') as follows:

(O_a) X ought_i to do a utilitarian action A in circumstances C , if and only if: there is a set $\alpha_h(A_1 \dots A_n)$ of α -alternatives open to X in C such that A belongs to $\alpha_h(A_1 \dots A_n)$ and for any member Z of this set different from A the value of kind K_i that X would bring about by doing A is greater than the value of kind K_i that X would bring about by doing Z .

There may be several sets of α -alternatives containing a certain utilitarian disjunction in common. (The sets $\alpha_3(AB)$ and $\alpha_4(AB)$ illustrate this.) We do not yet have a precise idea of how utilitarian disjunctions are assigned values. It is, thus, at this juncture an open question whether or not a disjunction that exceeds in value all other members of a set α_h , also exceeds in value all other members of a different set α_k of the same $(A_1 \dots A_n)$ family. Hence, it is still open whether or not by (O_a) an action is obligatory_i because of its membership in a given set of α -alternatives, and non-obligatory_i because of its membership in another set of α -alternatives.

The other way of accommodating disjunctive obligations requires that we give up the idea of a set of utilitarian alternatives that are all incompatible by pairs. We consider just one set of utilitarian alternatives, namely, the set $\mu^k(A_1 \dots A_n)$ which is the union of all the sets $\alpha_h(A_1 \dots A_n)$ of alternatives actually open to the agent in his circumstances.

Let $\mu(A_1 \dots A_n)$ be the set of all utilitarian actions built from the unanalyzed actions A_1, \dots, A_n . We define: a set of utilitarian μ -alternatives is a non-empty subset $\mu^k(A_1 \dots A_n)$ of $\mu(A_1 \dots A_n)$, such that if A is in μ and B is in μ^k , then $A \vee B$ is in μ^k . The reason is simply that if an action A is open to Utilius, then $A \vee B$ is open also to Utilius. As in the case of α -alternatives, the members of μ which are not in μ^k are the utilitarian actions not open to the agent in some specifiable circumstances. Let A and B be members of μ^k . We define: A is a μ^k -alternative to B , if and only if A and B are incompatible. Clearly, μ -alternativeness is not transitive.

On some methods of measuring the value of disjunctions it may be troublesome to consider

tautological acts. Thus, in the case of α -alternatives we may not want to count as a set that generates obligations the set whose sole member is the tautological act $A_1 \& \dots \& A_{n-1} \& A_n \vee A_1 \& \dots \& A_{n-1} \& \sim A_n \vee A_1 \& \dots \& \sim A_{n-1} \& A_n \vee \dots \vee \sim A_1 \& \dots \& \sim A_{n-1} \& \sim A_n$. Likewise, in the case of μ -alternatives we may want to exclude that act from every set μ^k . This is, however, a peripheral matter, and we shall say no more about it here.

In the case of μ -alternatives we do not have to worry about actions being both obligatory_i and non-obligatory_i on account of their membership in different sets of alternatives. Nevertheless, we must still beware of traps that may neutralize the reply to our preliminary argument in Sect. I. We cannot simply say that a given utilitarian action A is obligatory_i if and only if it is the highest-valued action in the relevant set μ^k . This would involve comparing compatible actions, which we are committed to avoid. We must, then, reinterpret (O) as follows:

- ($O\mu$) X ought_i to do a utilitarian action A , in circumstances C , if and only if: X 's doing A would bring about more value of kind K_i than X 's doing any action μ -alternative to A in a set $\mu^k(A_1 \dots A_n)$ of actions open to X in C .

It will be recalled that a set $\mu^k(A_1 \dots A_n)$ of open μ -alternatives is the union of all the sets $\alpha_i(A_1 \dots A_n)$ of open α -alternatives. Thus, we have as principle of the connection between α - and μ -alternativeness:

- ($\mu\alpha$) Every μ -alternative Z_A^μ to a utilitarian action A in a set $\mu^k(A_1 \dots A_n)$ of μ -alternatives open to an agent X in circumstances C is such that $V(A) > V(Z_A^\mu)$ [or, $V(A) \geq V(Z_A^\mu)$, or $V(A) < V(Z_A^\mu)$, or $V(A) \leq V(Z_A^\mu)$], if and only if: for every set $\alpha_i(A_1 \dots A_n)$ of α -alternatives open to X in C containing A , and for every alternative $Z_A^{\alpha_i}$ in α_i different from A , $V(A) > V(Z_A^{\alpha_i})$ [or, $V(A) \geq V(Z_A^{\alpha_i})$, or $V(A) < V(Z_A^{\alpha_i})$, or $V(A) \leq V(Z_A^{\alpha_i})$, respectively].

Since no set of alternatives is empty, and ($\mu\alpha$) is a necessary truth, we have the following two special principles, which will be used later on.

- ($\mu\alpha 1$) "For every μ -alternative Z_A to a utilitarian action A in a set $\mu^k(A_1 \dots A_n)$ of alternatives open to an agent X in circumstances C , $V(A) > V(Z_A)$ " entails "There

is a set $\alpha_h(A_1 \dots A_n)$ of α -alternatives open to X in C such that A belongs to α_h and for every member Z_A of α_h different from A , $V(A) > V(Z_A)$."

- ($\mu\alpha 2$) Like ($\mu\alpha 1$), except for having " \geq " instead of " $>$."

IV. OPENNESS OF ALTERNATIVES

The concept of action is very complex, and here we cannot analyze it. Yet we must be clear on some of its crucial features. We have so far assumed that given certain contexts or circumstances, certain actions may be taken as unanalyzed wholes. This presupposes some methods or principles for individuating and counting actions. Since this is a presupposition common to all theories of obligation, we let utilitarianism make this assumption. This does not mean that the study of how actions are individuated will not furnish new objections against utilitarianism. It simply means that it is legitimate to make that assumption while investigating what happens to utilitarianism from the side of deontic logic.

An important feature of the concept of action is its generality. An action is a doing that can be realized in different ways, or events. We shall say that an action allows different *mechanics*, and *mechanisms*, of performance. For instance, Utilius can pay his electric bill by personally paying in cash at some office, or by mailing a check, or by ordering his bank to make a deposit to the account of the electric company. Similarly, Utilius can mail his check by personally depositing it into a mailbox, or by requesting someone else to deposit it for him, or by pressing a button that will activate a mechanism that picks up letters and deposits them in a given mailbox. In these cases the mechanics of an action A are other actions, which are means of doing A , together with certain circumstances. The circumstances are such that in them the event of the agent's performing A is identical, though only contingently, with the event of his doing the action which is the means to A . For instance, Utilius' mailing a certain letter is exactly identical with his depositing the letter into a mailbox—if the circumstances are normal: the mailbox is in the route of a mailtruck, the truck does its usual round, etc. Likewise, Utilius pressing the button that activates his mailing mechanism is, if everything works as expected, identical with his mailing the letter.

Now, the circumstances may be such that

Utilius can bring about a disjunctive action $A \vee B$ either by bringing about A directly, or by bringing about B directly, or by bringing about the activation of a process that will end up with the realization of A or with the realization of B . These are different mechanics of Utilius doing $A \vee B$. In the first case he can bring about and choose to bring about A , in the second case he can bring about and choose to bring about B , and, hence, in both cases he can bring about and choose to bring about $A \vee B$. In the third case, however, Utilius can bring about A , if A is what eventuates, but he cannot choose to bring about A ; similarly, he can bring about B , if B is what his mechanism realizes, but he cannot choose to bring about B . In the third case Utilius can choose *merely* to bring about $A \vee B$, leaving it to nature, so to speak, to choose the specific disjunct. But if A is realized, when he activates his mechanism that will realize A or B , Utilius may be just as responsible for A as if he had himself chosen A .

Not everything an agent can bring about is, thus, something that he can choose to bring about. The courses of action which he can choose are precisely the courses of action open to him. Thus, not every action Utilius can bring about is always open to him. For an action to be open to him it is required that the action be in his power in the sense that he can actually choose to bring it about: it requires definite mechanisms connecting it to his capacity to choose.

One way of making clear the logic of the openness of utilitarian alternatives is to consider Utilius in circumstances such that whatever course of action he takes he will bring it about by pressing a button in an *action machine*. An action machine is simply a gadget having one button for each μ -alternative of a set $\mu(A_1 \dots A_n)$, where if the action A corresponding to button $b(A)$ entails the action B corresponding to button $b(B)$, then depressing $b(A)$ depresses $b(B)$. This physical connection between $b(A)$ and $b(B)$ is nothing but the physical realization of the logical connection between A and B . It is required for the machine to be an action machine, rather than something else. Now, given an action machine M with a button $b(A)$ we say that action A is *open* to agent X in the circumstances C corresponding to M if and only if $b(A)$ is in operating condition. Clearly, Utilius may be in circumstances such that his action machine M has $b(A \vee B)$ working while $b(A)$ and $b(B)$ do not work. Thus, we have the following important principles of openness:

(OA.1) If action A is open to agent X in circumstances C , and X 's doing A entails X 's doing B , then B is also open to X in C .

(OA.2) In some circumstances C , X 's doing A entails X 's doing B , yet B may be open to X in C while A is not open to X in C .

Let $\mu^k(A_1 \dots A_n)$ be the set of all alternative, open to Utilius in circumstances C . Let A be a member of μ^k . We shall say that A is a *principal alternative open to Utilius* (abbreviated *pua*) if and only if there are no members B and C of μ^k such that X 's doing $B \vee C$ is equivalent to X 's doing A . Clearly, given any set of $\mu^k(A_1 \dots A_n)$ of open alternatives, we can always produce the corresponding set $\mu^k_0(A_1 \dots A_n)$ of *principal open* alternatives by subtracting from μ^k every disjunction equivalent to the disjunction of two other members of μ^k . Naturally, all sets of α_k -alternatives we are interested in are sets of principal alternatives.

V. VALUE THEORY

Here we shall not be able to engage in a thorough study of value theory. The field is still in a state of fluidity, as can be seen from the illuminating discussion in Rescher [14]. Hallden [15], von Wright [2], and Chisholm-Sosa [17] are other worth-while contributions. But the most important work since Hallden's excellent monograph is Hansson [18]. For the allied area of decision theory, Jeffrey [19] is very promising.

Our task is the modest one of shedding some light on the implications for value theory of the several forms of act-utilitarianism. Our problem is well determined by the questions that our discussion of α - and μ -alternativeness left open: whether contradictions arise from (O α) or (O μ) on account of the methods for assigning values to utilitarian disjunctions. There is an argument in value theory as to whether, as was early suggested in [12], comparisons of value are primarily made only in terms of μ -alternatives, i.e., of utilitarian actions in the sense of this paper. But this argument does not affect our discussion here. Those who, like Hansson, directly compare the value of all propositions, regardless of their logical form, formulate principles that apply, *a fortiori*, to our utilitarian α - and μ -alternatives.

The issue whether values are assigned to propositions, or states of affairs, or facts can, too, be bypassed here. When we speak of the value of

an action, we must be understood as speaking of the value assigned to the appropriate entity: proposition, state of affairs, or fact to the effect that the agent has performed the action in question.

Regardless, then, of how in general one assigns values to actions, by assigning values to propositions, states of affairs, or facts, the following principles must hold, if utilitarianism is true:

- (V1) Utilitarian disjunctions must be assigned values, or must enter in comparisons of values.
- (V2) The value of a utilitarian disjunction cannot exceed the value of its highest-valued disjunct.
- (V3) The value of a utilitarian disjunction cannot be exceeded by the value of its lowest-valued disjunct.
- (V4) Von Wright's principle (VW) does not hold in general.

The truth of (V2) and (V3) can be easily gleaned from these facts: basic utilitarian actions (*bua*s) are complete possible states of the universe with respect to the non-analyzed primary actions A_1, A_2, \dots, A_n under consideration; they are mutually incompatible by pairs, so that the whole value accruing to the universe on account of the actions under consideration, i.e., the members of $\mu^k(A_1 A_2 \dots A_n)$, is exactly the value of one of the *bua*s in question; the value accruing to the universe from the performance of utilitarian disjunction is exactly the value of just one of the disjuncts of that disjunction.

Some hold (as Rescher does in [14], pp. 41 and 55) that the value to be assigned to a basic utilitarian action is simply the sum of the "raw" values (to use Rescher's term) of each non-analyzed primary conjunct. We do not hold this view. We are convinced that G. E. Moore was perfectly justified in defending (in [20] and [21]), what he called the *principle of organic unities*. This principle allows that the value of a conjunction be more, or be less, than the sum of the "raw" values of each conjunct, i.e., the value assigned in isolation to each conjunct. Nevertheless, here we shall not argue in support of the principle of organic unities. Our ensuing arguments do not depend at all on how the value of each *bua* is computed.⁴

It is often assumed that if the value assigned to a

proposition p is a real number n , then the value assigned to $\sim p$ is the number $-n$. This assumption has been attacked by Chisholm and Sosa in [17]. They adopt, instead, the principle that if the value assigned to p is positive, then the value assigned to $\sim p$ is not positive, and if the value assigned to p is negative, then the value assigned to $\sim p$ is not negative. They deal only with the "raw" values of primary, non-analyzed propositions, taken in isolation, and keep silent about the values of conjunctions, disjunctions, quantifications, and other compounds. We are not convinced that the Chisholm-Sosa principle of unequal numerical parity for the values of p and $\sim p$ is correct, even if restricted to really or metaphysically primary propositions (whatever they may be) in isolation. But we shall not attack it here. Our ensuing reasonings depend neither on its truth nor on its falsehood.

The preceding clarifies our assumptions, or, rather, lack of assumptions, concerning how non-analyzed propositions, their negations, and their conjunctions are assigned value measures. We can now proceed to disjunctions. Evidently, principles (V1)–(V4) above leave open three types of method for valuing disjunctions:

- (a) The value of a utilitarian disjunction is the same as the value of its higher-valued disjunct
- (b) The value of a utilitarian disjunction is the same as the value of its disjuncts, if these have the same value; otherwise the disjunction has a value that lies properly in between the values of its disjuncts
- (c) The value of a disjunction is the same as the value of its lower-valued disjuncts.

In the case of (b) we do not actually have a method, but only the limits of possible methods. There is an infinity of such possible methods, and there is no general reason to prefer one to another. Perhaps the most plausible methods of type (b) are: (i) the one proposed in earnest by Rescher in [14] that equates the value of a utilitarian disjunction with the arithmetical mean of the values of all its utilitarian disjuncts (*bua*s); and (ii) the probability measurement proposed by Jeffrey [19], p. 70, which has Rescher's as a special case, namely, the case in which all *bua*s are equiprobable.

⁴ In [26], chap. 2, Rescher gives very good reasons for rejecting the principle that in general the value of a conjunction is the sum of the values of the conjuncts. As he points out, of two distributions of the very same goods among the same people, one can be unjust, and thus of less value than the other.

VI. OUGHT-UTILITARIANISMS

We have three methods, (a)–(c), for measuring the value of utilitarian action, and two, α and μ , concepts of alternative utilitarian action. Thus, we have six interpretations of utilitarian principle (O). We shall refer to them with expressions of the form " $O(x, y)$," where ' x ' is ' α ' or ' μ ,' and ' y ' is ' a ,' ' b ,' or ' c ':

- $O(\alpha, a)$: ($O\alpha$) on the assumption that values are measured by method (a);
- $O(\alpha, b)$: ($O\alpha$) on the assumption that values are measured by same method of type (b);
- $O(\alpha, c)$: ($O\alpha$) on the assumption that values are measured by method (c);
- $O(\mu, a)$: ($O\mu$) on the assumption that values are measured by method (a);
- $O(\mu, b)$: ($O\mu$) on the assumption that values are measured by same method of type (b);
- $O(\mu, c)$: ($O\mu$) on the assumption that values are measured by method (c).

Each of these six principles characterize a variety of *simple ought-utilitarianism*. Each of these varieties connects oughtness to value by means of its characteristic $O(x, y)$ -principle, and then it deals with the rightness of actions by means of principle (DD1) in Sect. II above, namely: "An action is right _{i} if and only if it is not the case that it ought _{i} not to be done." The process is, of course, reversible. One can first connect rightness to value and then oughtness by means of the dual of (DD1). This reversal results in six views that we shall call *simple right-utilitarianisms*. We shall examine these views in the next section. (The terms "simple utilitarianism" and "general utilitarianisms" are taken from Lyons [22], pp. 25 ff.)

Our question is now multiple. Do these six ought-utilitarianisms pass the test of preserving the deontic entailments? For a better understanding of the logic of utilitarianism we shall break down each biconditional $O(x, y)$ -principle into its conditional parts. We shall call the 12 propositions that result *simple utilitarian ought-theses*.

For convenience of reference we shall state the utilitarian theses using the following abbreviations: " $V(\dots)$ " means, as before, "the value of kind K_i brought about by agent X performing the action \dots , which is open to him in circumstances C "; ' Z_A ' means "action Z alternative to A ,

different from A , also open to agent X in circumstances C ." Thus, all the actions discussed in the sequel are open to the agent in his circumstances. We shall consider our only agent to be Utilius and will omit explicit references to him and to his circumstances. Our twelve utilitarian theses are, then, obtained by the six xy -interpretations of each of the following two schemata:

- $O(x, y, N, g)$. If A is obligatory _{i} , then for every $Z_A: V(A) > V(Z_A)$.
- $O(x, y, S, g)$. If for every $Z_A: V(A) > V(Z_A)$, then A is obligatory _{i} .

In $O(x, y, N, g)$ the letter ' N ' indicates that the simple utilitarian condition of producing greater value is a *necessary* condition of obligatoriness. In $O(x, y, S, g)$ ' S ' indicates that the simple utilitarian condition is *sufficient* for obligatoriness. The letter ' g ' indicates that we are dealing with the relationship greater—than, i.e., $>$.

Actually, we have 24 ought-theses. The other 12 theses are like the 12 just introduced, except that the sentences formulating them have ' \geq ' instead of ' $>$ '. (Clearly, putting ' $<$ ' or ' \leq ' instead of ' $>$ ' is contrary to the spirit of utilitarianism which is concerned with maximizing value; hence, there is no point at all in considering the resulting theses.) Letting z be either ' N ' or ' S ,' we shall refer to theses involving the relation \geq rather than $>$, as the theses $O(x, y, z, ge)$, where " ge " stands for "greater than or equal to."

Fortunately, we can dispense with six theses at once. The six $O(x, y, S, e)$ -theses lead to contradictions. Let two incompatible actions A and B have the same value, as they may very well do. Then by whatever $O(x, y, S, ge)$ -thesis is adopted, Utilius is obligated _{i} to do A , and he is also obligated _{i} to do B . Consider the latter obligation. Utilius doing B entails Utilius not doing A , hence, by deontic principle (DR1) above in Sect. II, Utilius is obligated _{i} to abstain from doing A . By (DD1) and (O2) above, Utilius is *not* obligated _{i} to do A .

The remaining six $O(x, y, N, ge)$ -theses must be examined independently of each other. We have on the table, then, 18 utilitarian ought-theses.

Before proceeding to the examination of these theses let us grope for feasible shortcuts. To this aim, recall principle ($\mu\alpha$) and its derivatives, at the end of Sect. III. Principle ($\mu\alpha.1$) is roughly this: "For every Z_A alternative in μ^k : $V(A) > V(Z_A)$ " entails "There is a set α_h of α -alternatives containing A such that for every Z_A in α_h : $V(A) >$

$V(Z_A)$ ". That is, an ought-making $>-\mu$ -utilitarian condition entails its corresponding ought-making $>-\alpha$ -utilitarian condition. Hence, by $(\mu\alpha.1)$:

- I. If $O(\mu, y, N, g)$ holds, then its corresponding $O(\alpha, y, N, g)$ holds;
- II. If $O(\alpha, y, S, g)$ holds, then its corresponding $O(\mu, y, S, g)$ holds.

Principles I and II are very helpful, if they are, of course, employed in the right way. Given our preliminary argument in Sect. I above, our hunch is that some of the 12 theses related by Principles I and II are false. Guided by this hunch we shall proceed to examine the consequents of I and II, so that any falsity to be found will spill over the corresponding antecedents.

By principle $(\mu\alpha.2)$ in Sect. III above the remaining six $O(x, y, N, ge)$ -theses are related as follows:

- I. If $O(\mu, y, N, ge)$ holds, then its corresponding $O(\alpha, y, N, ge)$ holds.

Obviously,

- III. If $O(x, y, N, g)$ holds, then $O(x, y, N, ge)$ holds.

$O(\alpha, c, N, ge)$ is false. Let Utilius be obligated_i to do some action A . Then by $O(\alpha, y, N, ge)$ there is a set α_k of α -alternatives open to Utilius in some given circumstances, and $V(A) > V(Z_A)$ or $V(A) = V(Z_A)$ for every Z_A in α_k . Often such a set α_k has more than two members. Let it have as member action B , such that $V(A) > V(B) > V(\text{any other action in } \alpha_k)$. Let L be the basic utilitarian action with the lowest value. Now, on the one hand, by method (c): $V(A \vee L) = V(L)$, and $V(B) > V(A \vee L)$; hence every μ -alternative to $A \vee L$ is higher in value than $A \vee L$. Hence, in no set α_k does $A \vee L$ exceed or equal in value any of its alternatives. By $O(\alpha, y, N, ge)$ Utilius is, then, *not* obligated_i to do $(A \vee L)$. On the other hand, Utilius' doing A entails Utilius' doing $(A \vee L)$. Hence, by

principle (DR1) in Sect. II above, Utilius is obligated_i to do $(A \vee L)$. Therefore, $o(\alpha, c, N, ge)$ fails the test of deontic logic.

$O(\alpha, b, N, ge)$ must be rejected. Let Utilius be obligated_i to do a basic utilitarian action A . By $O(\alpha, y, N, ge)$ there is a set α_k of α -alternatives containing A , and $V(A) = V(Z_A)$ or $V(A) > V(Z_A)$ for every Z_A in α_k . Consider a case, which must undoubtedly be possible, in which $\alpha_k = \alpha_1$. Let α_1 be ordered as follows:

$$(o_1) \quad V(A) > V(B) > V(C_1) \geq \dots \geq V(C_r).$$

Since X 's doing A entails X 's doing $(A \vee C_1 \vee \dots \vee C_r)$, by (DR1) Utilius is obligated_i to do $(A \vee C_1 \vee \dots \vee C_r)$. Let $V(A \vee C_1 \vee \dots \vee C_r) = w$. By methods of type (b), w must lie somewhere between $V(A)$ and $V(C_r)$. From (o_1) , $V(B)$ also lies between $V(A)$ and $V(C_r)$. Furthermore, $V(B)$ is independent of the method of measuring utilitarian disjunctions. Thus, in many cases of the sort we are considering $V(B) > w$. Obviously, there is no other set of α -alternatives containing $A \vee C_1 \vee \dots \vee C_r$, than the pair which also contains B . Hence, there is no set of α -alternatives in which $A \vee C_1 \vee \dots \vee C_r$ exceeds or equals in value all of its co-alternatives. Hence, by $O(\alpha, b, N, ge)$ Utilius is not obligated_i to do $(A \vee C_1 \vee \dots \vee C_r)$.

The assumptions involved in the preceding discussion are of such generality that examples such as those described must be possible. For the sake of concreteness consider Rescher's method, which assigns to a utilitarian disjunction the arithmetical mean of the values of all its disjuncts, which are *bwas*. For greater emphasis suppose that we calculate the values of utilitarian disjunctions from the very beginning in the most stringent conditions, i.e., letting $V(\sim p) = -V(p)$ and $V(p \& q) = V(p) + V(q)$. With these narrow restrictions, consider the general case, where $m > n > 1$; and, hence, $m - n > -1/3(m - n)$;

Case No. 1 ($C \neq 1$)

Raw Values	$\alpha_1(AB)$	$\alpha_t(AB)$
$A : +m$	$A \& B : +(m+n)$	$A \& B \vee \sim A \& \sim B \vee \sim A \& B : -1/3(m-n)$
$B : +n$	$A \& \sim B : +(m-n)$	$A \& \sim B : +(m-n)$
$A : -m$	$\sim A \& B : -(m-n)$	
$B : -n$	$\sim A \& \sim B : -(m+n)$	

Here action $A \& B$ is the action supposed to be obligatory_i. $A \& B$ does satisfy $O(\alpha, b, N, ge)$, since it has the highest value in $\alpha_1(AB)$. By (DR1),

then $A \& B \vee \sim A \& \sim B \vee \sim A \& B$ is also obligatory_i. But this action does not surpass or equal $A \& \sim B$ in value in the only set $\alpha^k(AB)$ of α -

alternatives containing both. Thus, this disjunction is both obligatory_i and non-obligatory_i.

Another way of bringing out the ailment of $O(a, b, N, ge)$ is this. In the case tabulated in (C#1) what leads to the contradiction, one might say, implausible though it is, is not $O(a, b, N, ge)$, but the assumption that Utilius is obligated_i to do $A \& B$. What the argument shows is, one might continue, simply that in cases such as the above no action can be obligatory_i: all fail to satisfy the deontic principles and $O(a, b, N, ge)$ together. This is, formally, a legitimate reply. Suppose, however, as it may very well be the case, even in the presence of the stringent conditions listed in (C#1), that $m=n$. Then $V(A \& B)=2n$; $V(A \& \sim B)=0$, $V(\sim A \& B)=0$, and $V(\sim A \& \sim B)=-2n$. We have that $V(A \& B \vee \sim A \& \sim B \vee \sim A \& B)$, equal to 0, is not greater than $V(A \& \sim B)$. Thus, by $O(a, b, N, ge)$ no action should, then, be obligatory_i, if consistency is to be maintained. Now, if n is very large, then doing $A \& B$ would bring about a great bonanza, doing $A \& \sim B$ or doing $\sim A \& B$ would bring about nothing valuable, while doing $\sim A \& \sim B$ would bring about a great catastrophe. In other words, the case would be one in which, given the very nature of utilitarianism, bringing about the great bonanza should be, even if not straightout obligatory_i, at least *capable* of being obligatory_i. $O(a, b, N, ge)$ sins, thus, against the very soul of utilitarianism. It must, therefore, be rejected.

This examination of $O(a, b, N, ge)$ suggests that a method of type (b) that assigns to a utilitarian

disjunction a value very close to the higher disjunct is a better prospect. The limiting case of methods (b) is, of course, method (a). Thus, $O(a, a, N, ge)$ is the best $O(x, y, N, ge)$ -thesis, and $O(a, a, N, g)$ the best $O(a, y, N, g)$ -thesis

$O(a, a, N, g)$ must, however, be rejected. It contravenes the very spirit of utilitarianism. Suppose that the set of all principal alternatives open to Utilius in some given circumstances are the following utilitarian actions: $A=B_1 \vee C_1 \vee \dots \vee C_r, B_2$, and B_3, \dots, B_s , where each B_i and C_i is a *bu*. Let $V(B_1)=V(B_2)>V(B_i)>0$, for $i=3, 4, \dots, s$; let $V(C_j)<0$, for $j=1, 2, \dots, r$. Hence, by method (a), $V(A)$ is not greater than the value of any alternative to A in any set a_h containing A ; similarly, $V(B_2)$ is not the highest in value in any set of a -alternatives. Thus by $O(a, a, N, g)$ in this case no action can be obligatory_i. Yet Utilius' doing B_2 would guarantee that the largest amount of value would come about. On the other hand, assuming equal probabilities for every *bu*, Utilius' doing A has one chance out of $r+1$ that that same amount of value would come about, but it has r chances out of $r+1$ that some disvalue would result. Were Utilius to perform some of the other actions he would bring about either lesser value, or effective disvalue. In brief, the case under consideration is one in which it would at least be possible for Utilius to be obligated_i to do B_2 , or even some other B -action.

Perhaps a concrete illustration will not be amiss. Again, for emphasis, we assume the most stringent rules for the computation of values:

Case No. 2 (C#2)

Raw Values	$a_1(ABC)$	$a'_h(ABC)$
$A : +200$	$D_1=A \& B \& C : +200.1$	$a=D_1$
$B : +.1$	$D_2=A \& B \& \sim C : +200.1$	$b=D_2 \vee D_5 \vee D_6 \vee D_7 \vee D_8$
$C : 0$	$D_3=A \& \sim B \& C : +199.9$	$c=D_3$
$\sim A : -200$	$D_4=A \& \sim B \& \sim C : +199.9$	$d=D_4$
$\sim B : -.1$	$D_5=\sim A \& B \& C : -199.9$	
$\sim C : 0$	$D_6=\sim A \& B \& \sim C : -199.9$	
	$D_7=\sim A \& \sim B \& C : -200.1$	
	$D_8=\sim A \& \sim B \& \sim C : -200.1$	

Hence the set $a_h'(ABC)$ is the minimal set of open a -alternatives in the sense that no shorter disjunction of $a_h(ABC)$ is open to Utilius. In the case of action b , Utilius can perform the disjunction, without being able to determine which disjunct will be in fact actualized. He can choose the

disjunction as a whole from the four members of $a_h'(ABC)$ —but nature, so to speak, alone chooses the determinate disjunct. (We may conceive of Utilius being in a position where he cannot help but pushing one of four buttons: one for a , one for b , and one for c , one for d . Or, perhaps, he is to

choose between pushing one of three buttons and doing nothing, which automatically amounts to his doing one of the four actions.)

The point is that according to $O(a, a, N, g)$ Utilius cannot have a utilitarian obligation in the case tabulated in (C#2). Yet, if the units of value are very large, then +200.1 units of value is a great bonanza, while -199.9 is a catastrophe. Nevertheless, $O(a, a, N, g)$ makes it impossible that bringing about a bonanza while at the same time a catastrophe is prevented be obligatory. This sharply contradicts the idea that greater good be produced or greater bad be avoided.

$O(a, a, N, ge)$ must be rejected too. Let the units of value be very large indeed. Let the effective choice for Utilius be between the principal actions b and c above, as described in (C#2). That is, the sets of a -alternatives open to Utilius have just two members: b or c and disjunctions built on either of these two. Clearly, in no set of a -alternatives does c even equal its co-alternative in value. Hence, by $O(a, a, N, ge)$ c cannot be obligatory. Yet the former involves the certainty of 199.9 units of value in contradistinction to both the high likelihood of at least 199.9 units of disvalue and a small chance of 200.1 units of value. The very ideal of utilitarianism requires that in such a case action c should be possibly obligatory. Hence, $O(a, a, N, ge)$ must be rejected because of its anti-utilitarian streak.

We must, then, reject all three $O(a, y, N, ge)$ -theses. By Principle I we must reject $O(\mu, y, N, ge)$ -theses, and, by Principle III, all six $O(x, y, N, g)$ -theses. In sum, *simple utilitarian conditions are not in general necessary conditions of obligatoriness.*

$O(\mu, c, S, g)$ must be rejected. Consider the simple case of two unanalyzed primary actions A and B . Let the basic utilitarian actions (*buas*) that they yield be ordered as follows:

$$(o2) \quad V(A \& B) = x > V(A \& \sim B) = y > V(\sim A \& \sim B) = z > V(\sim A \& B) = w.$$

Let the set $\mu^k(AB)$ of all alternatives open to

Utilius in his circumstances consist just of all two- and three-membered disjunctions of the *buas*. Thus, Utilius is *not* obligated_i to do $A \& B$. Now, the set of open μ -alternatives to $A \& B \vee A \& \sim B$ has just one member: $\sim A \& B \vee \sim A \& \sim B$. By method (c), $V(A \& B \vee A \& \sim B) = y$ and $V(\sim A \& B \vee \sim A \& \sim B) = w$. Hence, by $O(\mu, y, S, g)$ Utilius is obligated to do $(A \& B \vee A \& \sim B)$. Likewise, the only open μ -alternative to $A \& \sim B \vee \sim A \& B$ is $A \& B \vee \sim A \& \sim B$. The former's value is z , and the latter's is w . Hence, by $O(\mu, y, S, g)$ Utilius is obligated to do $(A \& \sim B \vee \sim A \& B)$, too. By deontic principle (O3) in Sect. II above, Utilius is, then, obligated_i to do $[(A \& B \vee A \& \sim B) \& (A \& \sim B \vee \sim A \& B)]$. The bracketed conjunction is equivalent to $A \& \sim B$. Hence, by (DR1), Utilius is obligated_i to do $A \& \sim B$, in contradiction with the previous result.

$O(\mu, b, S, g)$ must also be rejected. Consider anew the case of two unanalyzed actions A and B satisfying (o2) above. Consider a method of type (b) that yields $V(A \& B \vee A \& \sim B) = \phi(x, y)$; $V(A \& \sim B \vee \sim A \& B) = \phi(y, z)$; $V(\sim A \& B \vee \sim A \& \sim B) = \phi(x, y)$; $V(A \& \sim B \vee \sim A \& B) = \phi(y, z)$; $V(\sim A \& B \vee \sim A \& \sim B) = \phi(z, w)$; $V(A \& B \vee \sim A \& \sim B) = \phi(w, x)$. Here in general, $u > \phi(u, v) > v$, if $u > v$. The values of the utilitarian actions are independent of the method ϕ . Clearly, in general, there is nothing that prevents, say, the following relationships from obtaining: $\phi(x, y) > \phi(z, w)$ and $\phi(y, z) > \phi(x, w)$. In such a case, if, as in the preceding paragraph no basic utilitarian action is obligatory_i, we find, by the same reasoning, that Utilius is both obligated_i and not obligated_i to do $A \& \sim B$.

It may be objected that the argument shows that restrictions must be placed on the values assigned to negations and conjunctions. But this is no defense, even if we forget the objectionability of some such restrictions. To fix our attention, let us take up Rescher's averaging method and, again, assume the very stringent rules: $V(p \& q) = V(p) + V(q)$, and $V(\sim p) = -V(p)$:

Case No. 3 (C#3)

Raw values	$\alpha_1(AB)$	$\mu^k_0(AB)$: principal alternatives
$A : +100$	$A \& B : +40$	$D_1 = A \& B \vee A \& \sim B : +100$
$B : -60$	$A \& \sim B : +160$	$D_2 = A \& B \vee \sim A \& B : -60$
$\sim A : -100$	$\sim A \& B : -160$	$D_3 = A \& B \vee \sim A \& \sim B : 0$
$\sim B : +60$	$\sim A \& \sim B : -40$	$D_4 = A \& \sim B \vee \sim A \& B : 0$
		$D_5 = A \& \sim B \vee \sim A \& \sim B : +60$
		$D_6 = \sim A \& B \vee \sim A \& \sim B : -100$

Set $\mu^k(AB)$ is the set of the principal alternatives open to Utilius. Thus, Utilius is *not* obligated_i to perform any basic utilitarian action in $a_1(AB)$. As (C#3) shows, D_1 is higher in value than its only μ -alternative D_6 , and so is D_5 whose only alternative is D_2 . Hence, by a reasoning exactly like the one made in connection with $O(\mu, c, S, g)$, we find that Utilius *is* obligated_i to A & B . We must, then, reject $O(\mu, c, S, g)$.

Finally, we must also reject $O(\mu, a, S, g)$. To jump into an example right away consider (C#3). Measure the value of the members of $\mu^k(AB)$ by method (a). The result is:

$$V(D_1) = 160; V(D_2) = 40; V(D_3) = 40; V(D_4) = 160; V(D_5) = 160; \text{ and } V(D_6) = -40.$$

By $O(\mu, a, S, g)$, Utilius is obligated_i to do D_1 , and he is also obligated_i to do D_5 . By the same reasoning made in connection with $O(\mu, c, S, g)$, Utilius *is* obligated_i to do A & $\sim B$, even if he is *not* obligated_i to do A & $\sim B$ or any other *but*, because they are not open to him in his circumstances.

Summing up the last paragraphs, we must reject all three $O(\mu, y, S, g)$ -theses. By Principle II above, we must also reject the three $O(a, y, S, g)$ -theses. We have already rejected the six $O(x, y, S, g)$ -theses. In short, *simple utilitarian conditions are not in general sufficient conditions for obligatoriness*.

VII. RIGHT-UTILITARIANISMS

It may be thought that all the difficulties we have lined up before utilitarianism are due to the fact that so far we have tested only ought-utilitarianism. It may be thought that right-utilitarianisms rest on a stronger foundation. They all begin with a seemingly weaker connection between value and deontic properties:

- (R') It is right_i that X do action A in circumstances C , if and only if: X 's doing A in C would bring about at least as much good (value) of kind K_i as (any) other action open to X in C .

Naturally, (R') merely characterizes *naïve* simple right-utilitarianism. However, we shall not rehearse a preliminary argument against naïve right-utilitarianism analogous to the argument in Sect. I above. Inasmuch as (R') is a proposition about *all* actions, its truth (or validity) is not affected if we limit ourselves to utilitarian actions (as defined in Sect. III). We hasten, then, to discuss sophisticated simple right-utilitarianisms.

As in the case of proposition (O'') above, once, we consider our two concepts of alternative action the three methods for measuring the value of disjunctive utilitarian actions, the necessity and the sufficiency of the maximizing condition, and the two relations \geq and $>$, we must distinguish *24 simple utilitarian right-theses* hidden in (R'). On the one hand, we have the instances, or interpretations, of the following schemata:

$R(x, y, N, g)$. If A is right_i, then for every Z_A : $V(A) \geq V(Z_A)$; and

$R(x, y, S, g)$. If for every Z_A : $V(A) \geq V(Z_A)$, then A is right_i.

On the other hand, there are the twelve instances of the following schemata:

$R(x, y, N, g)$. If A is right_i, then for every Z_A : $V(A) > V(Z_A)$; and

$R(x, y, S, g)$. If for every Z_A : $V(A) > V(Z_A)$, then A is right_i.

Obviously each $R(x, y, N, g)$ -thesis entails its corresponding $R(x, y, S, g)$ -thesis. Now, recall deontic principle O2: " A is obligatory_i" entails " A is right_i." Hence, by O2, each $R(x, y, N, g)$ -thesis entails its corresponding $O(x, y, N, g)$ -thesis, namely: " $\text{If } A \text{ is obligatory}_i, \text{ then for every } Z_A: V(A) \geq V(Z_A)$." In Sect. VI above, we have rejected all of the $O(x, y, N, g)$ -theses. Hence we must reject all six $R(x, y, N, g)$ -theses, and, because of these, also all six $R(x, y, S, g)$ -theses.

Evidently, each of the $R(x, y, S, g)$ -theses entails its corresponding $R(x, y, S, g)$ -thesis. Hence, we must start by examining the $R(x, y, S, g)$ -theses.

Recall now principle (μa) and its derivatives. Principle ($\mu a.1$) is this: "For every μ -alternative Z_A in μ^k : $V(A) > V(Z_A)$ " entails "There is a set a_n of α -alternatives containing A such that for every Z_A in a_n : $V(A) > V(Z_A)$." That is, every right-making μ -utilitarian condition entails its corresponding right-making α -utilitarian condition. Thus we have the principle:

If a $R(a, y, S, g)$ -thesis holds, then the corresponding $R(\mu, y, S, g)$ -thesis holds.

So, we shall take up the three $R(\mu, y, S, g)$ -theses. For convenience, we shall discuss the contrapositives of these theses. Schematically they can be formulated as follows:

$W(\mu, y, N, g)$. If utilitarian action A is wrong_i, then there is a set μ^k such that $V(A) \leq V(Z_A)$.

$R(\mu, c, S, g)$ and $W(\mu, c, N, g)$ must be rejected. They are both contrary to the very spirit of utilitarianism, and counter-intuitive in themselves. Consider the

example tabulated in (C#2), framed under the usual stringent rules. Its crucial data are as follows, where each D_i is a *bua*:

Case No. 4 (C#4)

$$\begin{aligned} D_1 &: +200.1; & D_2 &: +200.1; & D_3 &: +199.9; \\ D_4 &: +199.9; & D_5 &: -199.9; & D_6 &: -199.9; \\ D_7 &: -200.1; & D_8 &: -200.1. \\ e &= D_1 \vee D_2 \vee D_3 \vee D_4 \vee D_7 \vee D_8 : -200.1; \\ f &= D_5 \vee D_6 : -199.9 \end{aligned}$$

Let the set μ^k_0 of principal alternatives open to Utilius in his given circumstances consist of e and f . Again, we may picture the situation as one of either pushing a button to perform e or pushing another button (or doing nothing) to perform f . By $W(\mu, c, N, ge)$, i.e., $R(\mu, c, S, g)$, e cannot be obligatory_i. If it were obligatory_i, its contradictory action, namely f , would not be right_i. But f is right by $R(\mu, c, S, g)$. Yet if Utilius does f , the definitely right_i action, he will *certainly* bring about 199.1 units of disvalue; whereas if he does e his chances of bringing about only .2 more units of disvalue are, if all *bua*s are equiprobable, only one half of his chances of producing about 200 units of positive value! This by itself is sufficient to show that it must be possible for Utilius to be obligated_i, on purely utilitarian considerations, to do action e . We must, therefore, reject $W(\mu, c, N, ge)$ and its equivalent $R(\mu, c, S, g)$.

$W(\mu, a, N, ge)$ and $R(\mu, a, S, g)$ must be rejected. The reasons for rejecting these theses are the same as the reasons for rejecting $W(\mu, c, N, ge)$. Consider (C#4) above; this time, however, we employ method (a) and let μ^k_0 contain g and h :

$$\begin{aligned} g &= D_1 \vee D_2 \vee D_5 \vee D_6 \vee D_7 \vee D_8 : 200.1; \\ h &= D_3 \vee D_4 : 199.9 \end{aligned}$$

By $R(\mu, a, S, g)$ it is right_i for Utilius to do g . Hence, h cannot be obligatory_i for Utilius. Yet, h guarantees 199.9 units of value, while g has two to one chances of producing 199.9 or 200.1 units of disvalue. If no other considerations were valid, and utilitarianism ran the day, it should be possible for Utilius to be obligated_i to do h . We must, therefore, reject $R(\mu, a, S, g)$ and $W(\mu, a, N, ge)$.

Let us examine $R(\mu, b, S, g)$. Methods of type (b) for measuring the value of utilitarian disjunctions yield values between those given by method (a) and those given by method (b). As long as the method is not specified there is no reason to suppose that one cannot always, as for the case of methods (a) and (c), find situations in which the action that guarantees a bonanza cannot be

obligatory_i. Obviously, if Moore's principle of organic unities is true, basic utilitarian actions are assigned values in any fashion whatever. Thus, Utilius may very well find himself in a situation in which the following obtains: his only two μ -alternatives are, for instance, $A \& B \vee A \& \sim B \vee \sim A \& B$ and $\sim A \& \sim B$; the value of the former is just one unit over the value of the latter, and the probability that one disjunct of the former, with large disvalue, may occur is very great. In such a case it should be possible for $\sim A \& \sim B$ to be obligatory_i, even though $R(\mu, b, S, g)$ would not allow it.

A most plausible method of type (b) is the arithmetical-mean method. This method together with the much-used stringent rules that $V(\sim p) = -V(p)$, and $V(p \& q) = V(p) + V(q)$, provide a strongest framework for $R(\mu, b, S, g)$. Yet even within this framework Utilius can encounter a situation in which: $V(A) = +500.5$ and $V(B) = +499.5$. Hence, $V(A \& B) = +1000$; $V(A \& \sim B) = +.5$, $V(\sim A \& B) = -1000$. By the averaging method, $V(A \& B \vee A \& \sim B \vee \sim A \& B) = +.33$, while $V(\sim A \& \sim B) = -1$. By $R(\mu, b, S, g)$ Utilius cannot be obligated_i to do $\sim A \& B$. But if the likelihood of $\sim A \& \sim B$ being realized is very high, it should be possible, on utilitarian grounds, that it be wrong_i for Utilius to do the disjunctive alternative. Indeed, even with a small probability the catastrophe that $\sim A \& \sim B$ would bring about may be too great that a conservative utilitarian should make $\sim A \& B$ obligatory.

Summing up, all 24 simple utilitarian right-principles must be rejected. *Simple utilitarian conditions are neither necessary nor sufficient for rightness.*

VIII. PROBABILISTIC UTILITARIANISM

We have already noted that when the basic utilitarian actions (*bua*s) with highest value are not open, probabilistic considerations are very relevant. Thus, it is quite natural to conceive of a probabilistic version of simple utilitarianism. We think it evident that a probabilistic computation of the values assigned to propositions (or states of affairs, or actions) will not satisfy the deontic principles of distribution. But it may be of some interest to discuss one of the most important and most recent theories of probabilistic ranking of the preference assigned to propositions, to wit, that of Jeffrey's in [19]. With the tremendous progress in the theory of decision, some philosophers tend to believe that the concepts developed in the theory

of rational decision provide analyses of the deontic concepts on which traditional moral philosophers have, with undue narrowness, concentrated their discussions.

In addition to the axioms of probability, the crucial axiom in Jeffrey's logic of preferability or desirability ranking is, naturally, his axiom for assigning a preferability value to a disjunction, [19], p. 70. Abbreviating "probability of" to "prob()" and "desirability of" to "des()," and letting 'X' and 'Y' range over propositions, his axiom is:

- (5-2) If $\text{prob}(X \& Y) = 0$ and $\text{prob}(X \vee Y) \neq 0$, then $\text{des}(X \vee Y) = [\text{prob}(X) \text{des}(X) + \text{prob}(Y) \text{des}(Y)] \div [\text{prob}(X) + \text{prob}(Y)]$

This is a very nice formula. It reduces to Rescher's very plausible formula when all the *buas* entering into a utilitarian disjunction have the same probability. In that case: $\text{prob } X = \text{prob } Y$, and:

$$(5-2.0) (\text{des}(X_1 \vee X_2 \vee \dots \vee X_r) = 1/r \text{ des } X_1 + \text{des } X_2 + \dots + \text{des } X_r).$$

Thus the difficulties raised above for all simple utilitarian theses that employ the method of arithmetical mean are difficulties for the combination of utilitarian theses and Jeffrey's method.⁵

But Jeffrey's method fails to help utilitarianism even in cases in which the *buas* under consideration differ in probability. Consider Jeffrey's own example on p. 77:

Case No. 5 (C#5)

actions	prob	des	prob \times des
$A \& B$.4	1	.4
$A \& \sim B$.3	2	.6
$\sim A \& B$.2	3	.6
$\sim A \& \sim B$.1	4	.4
$A \& B \vee A \& \sim B$.7	1.43	
$A \& B \vee \sim A \& B$.6	1.63	
$A \& B \vee \sim A \& \sim B$.5	1.60	
$A \& \sim B \vee \sim A \& B$.5	2.40	
$A \& \sim B \vee \sim A \& \sim B$.4	2.50	
$\sim A \& B \vee \sim A \& \sim B$.3	3.33	

It is obvious in general, and the chart confirms it, that we cannot equate "Action Z is obligatory" with "The proposition that Z has been performed has the greatest desirability in the whole set." We might continue to resort to μ -alternatives, which are better than α -alternatives. We might, then, say: an act Z is obligatory, if and only if the proposition that Z has been performed has more desirability than any proposition to the effect that a μ -alternative to Z has been performed. But this won't do. From column "des" of the above chart we could read " $\sim A \& \sim B$ is obligatory." This by (DR1) implies " $A \& B \vee \sim A \& \sim B$ is obligatory." But the proposition that $A \& B \vee \sim A \& \sim B$ has been performed has value 1.60, while the proposition that is μ -alternative $A \& \sim B \vee \sim A \& B$

has been performed has value 2.40. Hence, $A \& B \vee \sim A \& \sim B$ is not obligatory as well.

Undoubtedly, Jeffrey's contribution is of prime importance. His analysis of the following case is highly illuminating. A dinner guest is to provide the wine, but has forgotten whether chicken or beef will be served. His telephone is not working, and other relevant conditions are met. His problem is to decide whether to take white or red wine. Given his circumstances, in particular that there will be beef if and only if there will be no chicken, and that he will bring red wine if and only if he will not bring white wine, Jeffrey, [19], p. 74, suggests the following data, where 'W' stands for 'white wine' and 'C' for 'chicken':

⁵ This is exactly how Brandt conceives of classical utilitarianism: "... the gain is historical truth, for we shall be assessing a theory closer to what the great utilitarians, such as Bentham, actually held. We shall therefore, construe ... the thesis of act-utilitarianism as follows: 'If doing A has, among *all the things* X can do, the maximum net expectable utility, then it is X 's objective duty to do A .'" [(27), p. 382; my italics.] Note that because of the italicized phrase the view formulated by Brandt falls under the very preliminary argument presented in Sect. I of this essay.

Case No. 6 (C#6)

wine & meat	prob.	des	prob \times des
$W \& C$	$.75p$	1	$.75p$
$W \& \sim C$	$.25p$	-1	$-.25p$
$\sim W \& C$	$.25(1-p)$	0	0
$\sim W \& \sim C$	$.75(1-p)$	1	$.75(1-p)$
$W \& C \vee W \& \sim C$	p	.50	
$\sim W \& C \vee \sim W \& \sim C$	$1-p$.75	
$\sim C$			

Since W is equivalent to $W \& C \vee W \& \sim C$, then the value assigned to the guest's action of bringing white wine is .50. Since $\sim W$ is equivalent to $\sim W \& C \vee \sim W \& \sim C$, then the value of the guest's action of bringing red wine is .75. In this case that there is chicken or that there is beef is a circumstance, or a state of affairs not under the guest's control. Thus the only open alternatives are bringing white wine and bringing red wine. Undoubtedly, the guest *ought* (given the above data), to bring red wine, i.e., the guest ought_i to bring red wine, where 'i' stands for "given the above data."

Yet, returning to our theoretical problem, the very same data above shows that it is not true in general that the guest ought_i to do an action that has greater Jeffrey-desirability than its alter-

natives. By deontic principle (O₄) in Sect. II above, if both it is the case that p , and the guest ought_i to do A , then the guest ought_i to do $p \& A$. Suppose that it is the case that chicken is served, then the guest ought_i to bring red wine while chicken will be served. If greater Jeffrey-desirability were a necessary condition for oughtness_i, then $W \& C$ above would have gotten a greater value than each of its μ -alternatives. But it does not have more value than its μ -alternative $W \& C$.

By the way, consider the natural suggestion that the value of interest to utilitarianism is the *probable utility* of a utilitarian action, computed by the numerator of Jeffrey's (5-2). In the case of Jeffrey's example tabulated in (C#5), we have: [27]

Case No. 7 (C#7)

actions	prob-utility	alternative actions	prob-utility
$A \& B$.4	$A \& \sim B \vee \sim A \& B \vee \sim A \& \sim B$	1.6
$A \& \sim B$.6	$A \& B \vee \sim A \& B \vee \sim A \& \sim B$	1.4
$\sim A \& B$.6	$A \& B \vee A \& \sim B \vee \sim A \& \sim B$	1.4
$\sim A \& \sim B$.4	$A \& B \vee A \& \sim B \vee \sim A \& B$	1.6

Here each row corresponds to an exhaustive pair of both α - and μ -alternatives. Hence the following remark applies to both relations of alternativeness: Clearly, by (O _{α}) or (O _{μ}) all four three-membered disjunctions are obligatory_i!

IX. GENERAL UTILITARIANISM

To complete our survey of act-utilitarianism we must say something, even if briefly, about one more type of utilitarianism.

Given principles (DD₂) (relating wrongness to oughtness) and (DD₁) (relating rightness to oughtness), what David Lyons has called *general utilitarianism* is equivalent to the conjunction of the following theses:

V. If it is right_i for someone to do A , then for every Z_A : $V(\text{everybody does } A) \geq V(\text{everybody does } Z_A)$.

VI. If for every Z_A : $V(\text{everybody does } A) \geq V(\text{everybody does } Z_A)$, then it is right_i for someone to do A .

Plainly, there are the corresponding ought-principles making action A obligatory for someone, or for everybody, if (only if) everybody doing A has more value than everybody doing any alternative Z_A .

All these general utilitarian theses are susceptible to the types of maladies that ail their instantiations $R(x, y, z, ge)$ above. For one thing, these general utilitarian theses hold for the case in which there is just one agent in the universe. In such a case we

are precisely limited to their instantiations. It might be said that the general theses require that there be a plurality of agents. Indeed, it might be said that utilitarianism is not a theory about moral obligation, but about political or social obligation. This is both an important concession and an important reply.

Let n be the minimum number of agents required by general utilitarianism. Clearly, for each agent we can produce tables of values for his actions as we did when discussing alleged $R(x, y, z, ge)$ -principles. For simplicity, consider those cases, as may very well obtain, in which the same actions are open to all the n agents, and the same values accrue to them. In some such cases the values of conjunctions may be additive. Then the kinds of differential relationships that hold between the values of utilitarian actions may be the same as in the examples considered above. The same considerations are applicable throughout. The actual details are laborious, yet it is apparent that the same maladies infect the general utilitarian principles, if they are generalizations of simple utilitarian conditions.

X. A VIABLE ACT UTILITARIANISM

We have found fault with 48 sophisticated simple utilitarian theses. The faults found with the nine weakest and most plausible of them are as follows, where "C.S.U." is short for "contrary to the very spirit of utilitarianism":

- $O(a, c, N, ge)$ and $O(a, b, N, ge)$: C.S.U. in some cases in which the highest-valued *buas* are open to the agent;
- $O(a, a, N, ge)$: C.S.U. in some cases in which the highest-valued *buas* are not open to the agent;
- $O(\mu, y, S, g)$: lead to contradictions in some cases in which the highest-valued *buas* are not open to the agent;
- $R(\mu, y, S, ge)$: C.S.U. in some cases in which the highest-valued *buas* are not open to the agent.

A rehearsal of the preceding discussion reveals two general results:

Result No. 1. If all the *buas* with the highest value in a set $\mu(A_1 \dots A_n)$ are open to agent X in circumstances C , then $O(\mu, a, S, g)$ and $O(a, a, S, g)$ both conform to the spirit of utilitarianism, satisfy deontic logic and yield exactly the same class of obligatory actions.

Clearly, if one *bua*, say A , exceeds all other *buas* in value, then every disjunction having A as one disjunct will exceed every one of its alternatives. Hence, A and every disjunction having A as disjunct will be obligatory by $O(\mu, a, S, g)$ and $O(a, a, S, g)$, without violation of deontic logic. Let k *buas* B_1, \dots, B_k have the same value and exceed in value each one of the remaining *buas*. Then in every set of alternatives to each $B_j, j=1, \dots, k$, there is at least one alternative to B_j not exceeded by B_j in value, namely another highest-valued *bua*, or the disjunction $B_1 \vee \dots \vee B_{j-k} \vee B_{j+1} \vee \dots \vee B_k$. Hence no B_j , or disjunction of B 's with fewer than k disjuncts will exceed in value all of its alternatives. On the other hand, every disjunction containing all the above k B 's will exceed in value every one of its alternatives.

Result No. 2. If all the *buas* with the highest value in a set $\mu(A_1 \dots A_n)$ are open to agent X in circumstances C , then $O(\mu, b, S, g)$ both conforms to the spirit of utilitarianism and yields the same class of obligatory actions as $O(\mu, a, S, g)$; but $O(a, b, S, g)$ may lead to contradictions.

Evidently, by method (b) given *buas* A, B , and C , $V(A \vee C)$ can be smaller than $V(B)$, even though $V(A)$ is greater than $V(B)$. In such a case there may be a set a_k containing B and $A \vee C$ such that on it $O(a, b, S, g)$ makes B obligatory, and a set a_h containing B and A such that on it $O(a, b, S, g)$ makes A obligatory. Since A and B are both *buas*, by DR1 they cannot both be obligatory. On the other hand, if $V(A)$ is greater than $V(B)$, $O(\mu, b, S, g)$ will never make B obligatory. Thus, in the case of method (b) it is important that we always consider μ -alternatives, and not a -alternatives.

The above two results constitute a part of the solution to the problem of formulating a viable act-utilitarianism that passes the test of deontic logic. But we still have to find a formulation that covers the cases in which the *buas* with the highest value are not open to the agent. These cases are very perplexing. Consider, for instance, the following cases, where the name of each *bua* has been usurped by its value, and a and b are the only principal alternatives, or the ones with highest value, intuitively speaking:

- (A) $a = (200 \vee -100 \vee -100)$; $b = 200$.
- (B) $a = (200 \vee 100 \vee 100)$; $b = 100$.
- (C) $a = (1000 \vee 1 \vee -1000)$; $b = -1$.
- (D) $a = (-50m \vee 10)$; $b = (-50m \vee -25m \vee 10)$.

In case (A) method (a) makes both a and b of equal value, whereas intuitively b is superior and should possibly be obligatory. Thus here a method different from (a) should be preferred. In case (B) methods (a) and (b) work well, while method (c) does not, for clearly in case (B) action a should be preferred to action b . In case (C) method (a) may yield the wrong answer, for the units of disvalue may be so large that even with a small probability that -1000 units may be realized we ought not, on utilitarian grounds, run the risk of producing them. *This suggests that there is a limit to the amount of disvalue a disjunctive alternative should allow for.* In case (D) if the *buas* have exactly the same probability, Rescher's method and Jeffrey's formula yield: $V(a) = -25m + 5$, and $V(b) = 25m \times 3.3$. Thus according to this way of computing values $V(a)$ is higher than $V(b)$. This is fine if we are concerned with the production of as much good as is probable. But if we are concerned with the avoidance of as much evil as we can probably avoid, then action b should be preferred; clearly, in the case we are considering, if one performs a one opens up a 50 per cent chance of producing 50m units of disvalue, where if one performs b one only opens up a $33 \frac{1}{3}$ chance of producing 50 units of disvalue.

In general, it seems that the avoidance of disvalue and, consequently, the risks of large disvalue should have certain priority. I frankly do not know now what is the best utilitarian view, but I want to formulate a version of act utilitarianism that appears to do justice both to the test of deontic logic and to the several intuitive insights brought out in the preceding discussion. To attain some degree of precision, we need a few preliminary definitions. Again, we assume that *buas* are assigned value somehow, and we shall continue to consider each *bua* as a one-membered disjunction.

Def. 1. The *pure value* of a utilitarian disjunction is the value of its higher-valued disjunct.

From now on we shall consider every utilitarian disjunction D to be of the form $D_1 \vee D_2 \vee \dots \vee D_r \vee D_{r+1} \vee D_{r+1} \vee \dots \vee D_{r+s}$, where the first r disjuncts have negative value and the remaining s disjuncts have positive value. Again "prob (X)" is

short for "probability of X " and " $V(X)$ " is short for "the pure value of X ." Given a utilitarian action D we shall assume that with respect to it, $\text{prob}(D_1) + \text{prob}(D_2) + \dots + \text{prob}(D_{r+s}) = 1$.

Def. 2. The *total negative value* of a utilitarian action $D = N(D) = \text{prob}(D_1) \times V(D_1) + \text{prob}(D_2) \times V(D_2) + \dots + \text{prob}(D_r) \times V(D_r)$.

Def. 3. The *total positive value* of a utilitarian action $D = P(D) = \text{prob}(D_{r+1}) \times V(D_{r+1}) + \dots + \text{prob}(D_{r+s}) \times V(D_{r+s})$.

Def. 4. L_i is the negative number, which represents the *upper limit of absolute undesirability* for the kind i of value under consideration.

Def. 5. A utilitarian action D has *more relative value* of kind i than a utilitarian action D' , if and only if one of the following cases obtain: (1) D' has a disjunct D'_i , such that $V(D'_i) < L_i$, while D has no such disjunct; (2) both D and D' have one such disjunct or neither has one such disjunct, but $N(D) < N(D')$; (3) D and D' both have one such disjunct as described in (1) or neither has it, and $N(D) = N(D')$, but $P(D) > P(D')$.

Now we can formulate an act ought-utilitarianism that satisfies the requirements discussed in this paper (but of course should be tested for other requirements):

(AOU) X ought _{i} to do A in circumstances C , if and only if: (a) X 's doing A would bring about more pure value of kind _{i} than X 's doing any μ -alternative to A in a set $\mu^k(A_1 \dots A_n)$ of actions open to X in C , and μ^k contains the highest-valued *buas* of the corresponding set $\mu(A_1 \dots A_n)$; or (b) X 's doing A has more relative value than X 's doing any other μ -alternative in a set $\mu^{(k_0)} A_1 \dots A_n$ of principal actions open to X in C , and μ^{k_0} does not contain the highest-valued *buas* of the corresponding set $\mu(A_1 \dots A_n)$; or (c) the command " X , do A " is entailed by a command " X , do B ", B and A belong to the set $\mu^k(A_1 \dots A_n)$ of alternatives open to X in C , and X ought _{i} to B .⁶

Indiana University

Received October 24, 1968

⁶ This essay is a partial result of research in moral philosophy carried on during 1967-1968, while the author was a Guggenheim Fellow.

BIBLIOGRAPHY

- [1] W. D. Ross, *The Right and the Good* (Oxford, Clarendon Press, 1928).
- [2] G. H. von Wright, *The Logic of Preference* (Edinburgh, Edinburgh Press, 1963).
- [3] —, *Norm and Action* (London, Routledge and Kegan Paul, 1963).
- [4] H.-N., Castañeda, "Actions, Imperatives, and Obligations," *Aristotelian Society Proceedings*, 1967–68, pp. 25–48.
- [5] —, "The Logic of Change, Action and Norms," *The Journal of Philosophy*, vol. 62 (1965), pp. 333–344.
- [6] —, "A Note on Deontic Logic," *The Journal of Philosophy*, vol. 63 (1966), pp. 231–234.
- [7] —, "Acts, the Logic of Obligation, and Deontic Calculi," *Philosophical Studies*, vol. 19 (1968), pp. 13–26.
- [8] —, "Outline of a Theory on the General Structure of the Language of Action," *Theoria*, vol. 36 (1960), pp. 151–182.
- [9] —, "A Theory of Morality," *Philosophy and Phenomenological Research*, vol. 17 (1957), pp. 339–352.
- [10] —, "Imperatives, Decisions, and Oughts: A Logico-metaphysical Investigation" in H.-N. Castañeda and G. Nakhnikian (eds.), *Morality and the Language of Conduct* (Detroit, Wayne State University Press, 1963), pp. 219–299.
- [11] —, "Ethics and Logic: Stevensonian Emotivism Revisited," *The Journal of Philosophy*, vol. 64 (1967), pp. 671–683. "Errata," *ibid.*, p.
- [12] —, Review of [15] below, *Philosophy and Phenomenological Research*, vol. 19 (1958), p. 266.
- [13] D. Sidorski, "A Note on Three Criticisms of von Wright," *The Journal of Philosophy*, vol. 62 (1965), pp. 739–742.
- [14] N. Rescher, "Semantical Foundations for the Logic of Preference" in N. Rescher (ed.), *The Logic of Decision and Actions* (Pittsburgh, University of Pittsburgh, 1968), pp. 37–62.
- [15] S. Hallden, *On the Logic of 'Better'* (Uppsala, Library of *Theoria*, No. 2, 1957).
- [16] R. M. Chisholm, "Contrary-to-Duty Imperatives and Deontic Logic," *Analysis*, vol. 24 (1963), pp. 33–36.
- [17] — and E. Sosa, "On the Logic of Intrinsically Better," *American Philosophical Quarterly*, vol. 3 (1966), pp. 244–249.
- [18] B. Hansson, *Topics in the Theory of Preference Relations*, mimeographed essay by the Philosophy Department of the University of Lund, 1966.
- [19] R. C. Jeffrey, *The Logic of Decision* (New York, McGraw-Hill, 1965).
- [20] G. E. Moore, *Principia Ethica* (Cambridge, Great Britain, Cambridge University Press, 1903).
- [21] —, *Ethics* (London, Oxford University Press, 1912).
- [22] D. Lyons, *Forms and Limits of Utilitarianism* (Oxford, Clarendon Press, 1965).
- [23] W. Sellars, "Reflections on Contrary-to-Duty Imperatives," *Noûs*, vol. 1 (1967), pp. 303–344.
- [24] L. Aqvist, "Good Samaritans, Contrary-to-Duty Imperatives, and Epistemic Obligations," *Noûs*, vol. 1 (1967), pp. 361–379.
- [25] L. Powers, "Some Deontic Logicians," *Noûs*, vol. 1 (1967), pp. 381–400.
- [26] N. Rescher, *Distributive Justice* (New York, Bobbs-Merrill, 1966).
- [27] R. B. Brandt, *Ethical Theory* (Englewood Cliffs, N.J., Prentice Hall, Inc., 1959).

II. CONSEQUENTIALISM AND ITS COMPLEXITIES

MICHAEL STOCKER

HENRY Sidgwick and G. E. Moore can be read as holding that the rightness or wrongness of an act depends, in a conceptual way, solely upon the causal consequences of that act. J. S. Mill, on the other hand, can be read as holding that the rightness or wrongness of an act depends, in a conceptual way, solely upon the causal consequences of compelling a person to do that act or punishing him for not doing it.¹ Read this way, these philosophers in their different ways, and others in similar and still other ways, are *consequentialists*—holding that such act evaluations depend, in a conceptual way, solely upon the evaluation of the consequences of some act or acts. W. D. Ross and others rejected this view, holding that such act evaluations at least sometimes depend in a conceptual way on what is *intrinsic* to an act.

In this paper I shall attempt to examine and clarify the dispute over consequentialism. I shall not, however, be interested solely in historical disputes. I shall be interested also in the question of whether any consequentialist position is correct or even plausible.

At the outset, I wish to make it clear that throughout this study I am primarily interested in those consequentialist theories that present factors that (purportedly) *make* acts right, wrong, and so on; factors *because* of which or *in virtue* of which an act is right, wrong, and so on.² Nonetheless, many of my comments bear on other sorts of theories—e.g., those that present factors that are (purportedly) only necessary and sufficient conditions for an act's being right. For example, as I understand William Paley, his utilitarian theory is not, as, e.g., Mill's is, of the former sort. For Paley holds that an act is right just because God or some divine agent so designates it, but that God or the divine agent so designates all and only those acts that a utilitarian would hold to be right.

Given this requirement for consequentialism, I shall argue that it may be somewhat more reasonable to believe that there is no correct consequentialist theory than it is to believe that there is one. As interested as I am in showing this, I am at least equally interested in showing that this is the strongest justifiable conclusion that can be reached.

I. WHAT CONSEQUENTIALISM IS

A given evaluation *E* (e.g., right) of a given act is a *consequential evaluation* of that act just in case that act is *E* solely because some act or acts have certain consequences. An ethical theory about a class of act evaluations is a *consequentialist theory* just in case it holds that all of these evaluations are consequential evaluations. This understanding of a consequentialist theory allows, and is intended to allow, that there be many different, specific forms of consequentialism. For it leaves unanswered such questions as: (A) Which act evaluations are consequential evaluations? For example, are "right" and "wrong" the only consequential evaluations? (B) Which sort or sorts of consequences are relevant, or as I shall say *morally relevant*, for consequential evaluations? For example, are *hedonic* consequences the only morally relevant consequences? (C) What "function" of these morally relevant consequences determines or gives the consequential evaluation? For example, is *maximizing their overall value* the only plausible function? (D) How is the act (or acts) which has the consequences relative to which a given act is evaluated related to that act? For example, must it be the *very same* act?

No one of these questions, except perhaps (A), is answered in the same way by all historical consequentialists. Nor need all "possible consequentialists" answer any of them in the same way. Thus, no particular consequentialist theory is

¹ Henry Sidgwick, *The Methods of Ethics*, 7th ed. (London, 1963), passim, e.g., p. 411; G. E. Moore, *Ethics*, (Oxford, 1961), passim, and *Principia Ethica* (Cambridge, 1959), sect. 89, but see sect. 17; J. S. Mill, *Utilitarianism* (New York, 1948), ch. V, passim, esp. paragraphs 12 ff, cf. this paper sect. 7. Henceforth, by "consequence" I shall mean "causal consequence."

² Henceforth, unless something turns on the distinctions among such act evaluation notions as right, wrong, all right, obligatory, duty, ought, I shall use one to stand indifferently for them all.

the consequentialist theory. And therefore, showing that some particular consequentialist theory is mistaken is just that—showing that *some particular* consequentialist theory is mistaken. For these reasons I shall be concerned only with those considerations that bear against consequentialism as such, and not simply against particular consequentialist theories.

This concern is modified to some extent. For if all that could be said in support of consequentialism were that only a few evaluations of a few acts are consequential evaluations, consequentialism would hardly be controversial. (Ross, for example, was not interested in denying that some act evaluations are consequential evaluations.) Consequentialism is controversial when it is held that such act evaluations as right, wrong, permissible, one's duty, what one ought to do (or not do) are all consequential evaluations. And, thus, in this paper I shall consider only those consequentialist theories that hold that all such act evaluations are consequential evaluations. (Ross, for example, was interested in denying that such a consequentialist theory could be correct.)

II. THE ANTI-CONSEQUENTIALIST ARGUMENT

In this section I shall present what I take to be the historically and conceptually important anti-consequentialist argument. The argument is: (i) at least some acts are actually (i.e., not just *prima facie*) right because they are acts of promise-keeping; (ii) the fact that each such act is made actually right by being an act of promise-keeping is not dependent solely upon (considerations of) consequences; (iii) the fact that each such act is an act of promise-keeping is not dependent solely upon (considerations of) consequences; therefore (iv) not all evaluations of acts, nor even all evaluations of acts as actually right are consequential evaluations; and therefore (v) consequentialism is mistaken.³

When directed against consequentialism as

such, and not simply against the different consequentialisms of Bentham, Mill, Sidgwick, Moore, Rashdall, and Joseph, this anti-consequentialist argument has a number of gaps which can serve as places for consequentialist rejoinders. Meeting these rejoinders should show why it is difficult to declare with confidence either for or against consequentialism.

(i) is, I think, true beyond question. Therefore, the consequentialist must show either (ii) or (iii) mistaken. In what follows, I shall attempt to point out how the consequentialist might try to show this. To this end, I shall consider two non-exclusive consequentialist objections which I shall call (vi) the *direct* and (vii) the *indirect* objections to this anti-consequentialist argument. The direct objection, (vi), is that each of these acts of promise-keeping has consequences in virtue of which it is consequentially evaluable. The indirect objection, (vii), is that another act or other acts have consequences in virtue of which each of these acts of promise-keeping is consequentially evaluable.

III. INTRINSIC PROPERTIES AND CONSEQUENTIAL PROPERTIES

It is usually not denied that acts that fulfill (or fail to fulfill) e.g., the duty of nonmaleficence are consequentially evaluable.⁴ What is denied is that acts that fulfill (or fail to fulfill) e.g., the duties of reparation, gratitude, or promise-keeping—i.e., (Ross's) duties of *special obligation*—are consequentially evaluable. And I shall assume that if such an argument as presented by (i)–(v) works at all, it works only against acts that fulfill (or fail to fulfill) duties of special obligation.

The consequentialist might suggest, in the following way, that even acts that fulfill such duties are consequentially evaluable. He can note that an act cannot even be an act that discharges a duty of reparation or gratitude unless it produces benefit for someone.⁵ Similarly, it would be argued that promises are binding only if made

³ Two notes on this argument: (1) This argument, and especially (i), should not be taken as denying that *if* an act of promise-keeping had certain consequences (which it does not have) then it would not be right. But the absence of such potentially overriding conditions need not, I believe, be taken as even part of the ground of the rightness of all acts of promise-keeping that are right to do. (2) Henceforth, the context will be relied upon to distinguish between actual and merely *prima facie* rightness, wrongness, and so on.

⁴ At times Ross writes as if acts that discharge (or fail to discharge) duties of non-special obligation—e.g., of nonmaleficence—are consequentially evaluable; at times he denies this. Cf. *The Right and the Good* (Oxford, 1961), ch. ii, *passim*, esp. pp. 21, 25, and 47.

⁵ I shall defer the two following problems. (1) At times one can discharge (or fail to discharge) such duties simply by not doing (e.g., by abstaining from doing) some act. Thus an act that discharges (or fails to discharge) such a duty need not causally produce any benefit (or harm); rather it may only *avoid* producing harm (benefit). I discuss this under the rubric of

with the point of benefiting or satisfying some specified person(s); and if keeping the promise will not benefit or satisfy them, then the promise is not binding; and thus keeping it is not right, and not keeping it is not wrong.

These considerations do not appear to be conclusive. For example, Ross's anti-consequentialist position might be put this way. That an act of promise-keeping is right may require that it have consequences of a certain sort (and lack consequences of another sort); but not all the rightness of each act of promise-keeping is attributable to these or other consequences; at least some of the rightness must be attributed to something that is *intrinsic* to the act—viz., simply *being an act of promise-keeping*.

In the next several paragraphs I wish to assess the relevance for consequentialism of the claim that some morally relevant properties—e.g., *being an act of promise-keeping*—are intrinsic to acts. In this section I shall try to show (i) that it does not clearly follow that if a property is intrinsic to an act it is therefore not a consequential property of that act; and in the next section I shall try to show (ii) that even if a morally relevant property is an intrinsic and (perhaps thus) a non-consequential property of an act, consequentialism is not thereby shown false.

To show this, it will be necessary to sketch several views of intrinsic properties. Intuitively, I would say an intrinsic property of something is a property of that thing that is essential to its being what it is. Roughly speaking, it is a property that the thing has such that if it lacked that property it would not be what it is.⁶ Given this intuitive, and I believe common, view of intrinsic properties, we are, I think, committed to the view that all properties are intrinsic. For consider any property *p* of a given particular *Q*. One of the things *Q* is, is a *particular with property p*. But obviously it could not be this if it lacked *p*. But if all properties of particulars are intrinsic, and if at least some properties of acts are consequential properties, we cannot argue from the fact that a property is intrinsic to an act to the conclusion that the property is not a consequential property of that act.

Since the implications of the above view of

intrinsic properties are so troublesome, rather than asking whether or not a property is intrinsic to a given particular, I shall follow a well-worn path and ask only whether a property is intrinsic to a given particular *under a certain description*; that is, whether a property is intrinsic to a given particular *qua* or *as a such-and-such*. And I shall say (again roughly) that a property is intrinsic to something *qua* such-and-such just in case that thing could not be a such-and-such if it lacked that property, i.e., just in case that description could not be true of it if it lacked that property. Thus the question of whether some given property is intrinsic to a particular act is now posed by asking whether that property is intrinsic to that act under a given description.

One of the more important questions, then, is: What descriptions of that act should be adverted to? Here, however, we may be faced with a serious problem about the nature and individuation of acts. While it may be clear that such material particulars as chairs can be described in various ways, it is not at all clear that this is so for acts. Consider for example, "a chair my uncle likes," "a chair made by John's father," and "the heaviest piece of furniture in the room." All of these, I think, can be, and can easily be seen to be, descriptions of one and the same particular—that thing over there. Even though relative to these various descriptions different properties are intrinsic to the chair, they are all properties of the very same chair. But it is not clear that when philosophers speak of different descriptions of the same act they are speaking of various descriptions of the same act. When it is said that we can describe what Jones is doing as *reading a book*, or *pleasing his wife*, or *pleasing his wife by reading a book* it is not clear (to me at least) how many acts there are. It is not clear that each of these descriptions indicates a different aspect of the same act—as do the various descriptions of the chair—or, rather, whether they "give" different acts, different particulars, e.g., to be evaluated. In this regard, I find it more difficult to answer the question of how many things I did yesterday, or am doing now, than to answer the question of how many things there are in this room.

negative-consequentialism in Sect. VII. (2) This claim (as most of the claims in this paper) is cast in the objectivist mode. But perhaps one has discharged a debt of gratitude or reparation if one has tried to benefit the appropriate person, whether or not one succeeds. Simply for reasons of simplicity of exposition, I shall use the objectivist mode; my claims are easily translated into the subjectivist mode.

* ⁶ I say *roughly* since at least some might want to hold that there are some properties a particular necessarily has even though such properties are not intrinsic or essential to that particular—e.g., such properties as *being red* or *not being red*, or those properties necessitated in a synthetic *a priori* way by intrinsic properties.

When speaking of different descriptions of the same material particular, we have spatio-temporal coordinates as a touchstone of identity. But do we have any such touchstone, or any touchstone at all, of act identity? Do we have any, that is, besides the description or that which is given by the description? As various philosophers have argued recently, it seems plausible to suggest that the only, or the important and the authoritative touchstone of act identity is just the description of the act.⁷ But then, as they point out, this entails that different descriptions of what is supposed to be the same act, or at least many such descriptions, are not descriptions of the same act, but rather "give" or are of different acts.

Suppose that we can, or should, take different act descriptions as different descriptions of the same act. Then, as in regard to the first way of understanding intrinsic properties, just any property of an act will be intrinsic to the act, but now only relative to various descriptions of it. (To see that any property *p* of any act *Q* can be intrinsic to *Q* relative to some description, consider "an act that has *p*" as the relevant description of *Q*.) But if just any property of an act can be intrinsic to an act, and if it is intrinsic only relative to some descriptions, and if there are consequential properties of that act, it cannot follow that if a property is intrinsic to that act it is not a consequential property of that act. At most, it could be held that if a property is intrinsic to an act relative to a given description, then if it is a consequential property of that act, it is a consequence only relative to various other descriptions of the act.

If, as suggested above, we cannot take different, or all the different, descriptions of what is supposed to be one act to be descriptions of the same act, then we reach the following result. It is not that each property of an act is, under some description of that act, an intrinsic property of that act. Rather, each property of an act is an intrinsic property of some act(s) or other. Consider the act given by "Jones hitting a ball." Let us say that this act has

the consequential property of *being the cause of Smith's broken window*. Can that property be intrinsic to that act? According to the view under consideration, the answer is negative. Thus, it would be mistaken to claim that relative to "Jones hitting a ball, which breaks Smith's window" that property is intrinsic to the act given by "Jones hitting a ball." For these are descriptions, according to this present view, of different acts. In producing a description relative to which the property is intrinsic, we get or give a different act. Nonetheless, relative to the act which is given by "Jones hitting a ball, which breaks Smith's window" the property of *being the cause of Smith's broken window* is intrinsic.

Here the position about the relationship between a property's being intrinsic and consequential is that even though a property is an intrinsic and (perhaps thus) a non-consequential property of one act, it is *possible* for that same property to be a consequential and (perhaps thus) a non-intrinsic property of another act. Here, of course, there is no contradiction between holding that there are both intrinsic and consequential properties of acts and also that no property can be an intrinsic and a consequential property of the same act.⁸

IV. SOME IMPLICATIONS FOR THE ANTI-CONSEQUENTIALIST ARGUMENT

If we accept the view that no property can be an intrinsic and a consequential property of the same act or if we accept the view that different descriptions give different acts, then in examining the consequentialist rejoinders to the anti-consequentialist argument, we shall have to switch at least some of our attention from the direct objection, (vi), to the indirect objection, (vii). For consider a consequentialism that holds that the only morally relevant properties of acts are hedonic ones—pleasure and pain, for short. On its face, it might appear that this consequentialism would be shown mistaken if we could exhibit at

⁷ David Sachs appears to suggest something like this in "A Few Morals About Acts," *The Philosophical Review*, vol. 75 (1966), pp. 91–92. Arthur Cody in "Can A Single Action Have Many Descriptions?," *Inquiry*, vol. 10 (1967), pp. 164–180, Alvin Goldman and Robert Ware (both in unpublished work) clearly hold this position. I am indebted to their work and should like to express my thanks to Goldman and Ware for showing me their work on this topic and for discussing it with me. For a contrary position on act individuation see Donald Davidson's article in *The Logic of Decision and Action*, ed. by N. Rescher (Pittsburgh, 1968).

⁸ It might be objected that some of the "acts" or descriptions we shall have to cite to make this argument work are not really acts or act-descriptions. But as Sachs (*op. cit.*) and White (introduction to *The Philosophy of Action* [Oxford, 1968], point out, when philosophers speak of acts, they do not use "act" in ordinary or usual ways. Rather they collect under that term most, if not everything, that we can be said to do or not do. And my "acts" clearly are things we do or do not do. I am indebted to Robert Ware for discussing this with me and for showing me his unpublished work on the subject.

least one act that is wrong to do, but that lacks any consequential, hedonic qualities. For example, consider the act of *displeasing one's wife*, or of *displeasing one's wife by reading a book*. The hedonic properties that make these acts wrong are intrinsic, not consequential.

Nonetheless, such considerations are not strong enough to show hedonic consequentialism mistaken. The consequentialist can simply shift from the direct rejoinder, (vi), to the indirect rejoinder, (vii), and argue that the relevant hedonic property, (*the wife's*) *being displeased*, is a consequential property of another act—viz., *reading a book*. He can further point out that this act is importantly related to the other acts; and, in fact, the other acts' having that hedonic property intrinsically is a consequence of the act of reading a book. That is, the other acts have it intrinsically because *reading a book* has it consequentially.

In short, to show consequentialism mistaken, it is not sufficient to exhibit an act that is made wrong by properties intrinsic to it. For given some descriptions of an act, or the proper choice of acts, just any property of an act is intrinsic. Rather, as (vii) is calculated to point out, one must also show that another act's or other acts' having that property consequentially is not a sufficient condition for the act's having it intrinsically.⁹

V. ACTS AS CAUSES OF ACTS

Having now prepared the way, I shall begin to examine the consequentialist rejoinders. In this section I shall examine part of the indirect objection, (vii), that even though a given act does not itself have consequences in virtue of which it is consequentially evaluable, another act or other acts do have consequences in virtue of which that act is consequentially evaluable. In Sect. VI, I shall turn to the direct objection, (vi); and in Sect. VII, I shall consider other indirect objections. I have divided (vii) into two parts since the part dealt with in this section, but not the part dealt with in Sect. VII, leads naturally into (vi), the subject of Sect. VI. In this present examination of (vii), as part of the argument to defend consequentialism, it is assumed that different descriptions of acts, or at least those different des-

criptions under discussion in this section, give different acts. In Sect. VI, it will be shown how consequentialism can be defended in a way that avoids this view of act individuation; however, the arguments presented in this section and in Sect. VI are otherwise essentially the same.

As was shown in the previous section, to defeat consequentialism it is not sufficient to exhibit an act that is made wrong by properties intrinsic to that act. It must also be shown that no other act (or acts) has consequences in virtue of which that act is consequentially evaluable. Since it should be agreed, I believe, that an act can be right because it is an act of promise-keeping, where the property of *being an act of promise-keeping* is intrinsic to that act, the consequentialist must contend that an act's having that property intrinsically is due to some act's or acts' having it consequentially. One way to do this is to show that the act of keeping a promise is itself a consequence of another act or acts.

Thus, in this section, I shall consider the consequentialist claim that acts of promise-keeping are consequences of acts (and *mutatis mutandis* that acts of promise breaking, and other acts that discharge or violate other duties of special obligation are themselves consequences of acts). This part of the indirect rejoinder is, thus, a criticism of premiss (iii) of the anti-consequentialist argument: viz., the fact that each such act is an act of promise-breaking is not solely dependent upon (considerations of) consequences. Here the consequentialist argument runs as follows: If an act of promise-keeping is itself a consequence of some other act, then the consequences of that other act are what make the act of promise-keeping right. Here the morally relevant consequence of the other act is just the act of promise-keeping.

More fully, the argument is as follows. We often do one act *by* doing another. This "by" can stand for any number of relations—e.g., causal, conventional, logical, coextensional, and perhaps others. *Causally*, I broke the window by hitting it. *Conventionally*, I checkmated him by moving my rook to . . . *Logically*, I reached a true conclusion by giving a sound argument. *Coextensionally*, (merely by coincidence) I did what he most wanted me to do at that moment by jumping up and down.¹⁰

⁹ A similar position is suggested by Jan Narveson in "Utilitarianism and Formalism," *The Australasian Journal of Philosophy* vol. 43 (1965), pp. 58-72.

¹⁰ I am indebted to Alvin Goldman for discussing the role of such "by's" with me, and for showing me his unpublished work on the subject. His suggestions led me to think of the following consequentialist rejoinder. I do not know, however, if he agrees with the use to which I put his ideas.

The consequentialist rejoinder now under consideration is that an act of promise-keeping is done *by* doing another act or other acts, where this "by" is a causal "by," and thus the act of promise-keeping is, itself, a consequence of the other act(s). Another way to put this is that the property of a *promise's being kept* is a consequential property of the other act(s). The final part of the claim is that this consequential property is a morally relevant property.

A person can causally bring about a lantern's being lighted by flipping a switch. (a) He lights the lantern by flipping the switch. Given certain circumstances, the lantern's being on or being turned on is, or counts as, a signal. (b) The signal is given by the lantern's being on or being turned on. The "by" of (a) is a causal one, and the "by" of (b) is a conventional or coextensional one. Can we therefore say that the *signal's being given* is a causal consequence, or a consequential property of the flipping of the switch? It seems to me that the answer is affirmative. After all, we can say (c) he gives the signal by flipping the switch. It seems to me that the flipping of the switch, which is not the signal, causes the signal to be given.

The application of this to promises should be clear. Suppose that Jones has made a binding promise to return Smith's book. Let us say that delivering the book to Smith is returning it to him, or that by which it is returned to him. If Jones causes it to be delivered—e.g., by shooting it out of a cannon, or dropping it in a mail box, or hiring a messenger—does he also cause it to be returned? Again, it appears to me that the answer is Yes. But returning the book is what he promised to do. It would appear that he therefore causes to take place what he promised to do or that he would make happen. Can we say that he therefore caused the promise to be kept, or that *keeping the promise* is a consequential property of what he did, or of one of the things he did? Again, it seems to me that an affirmative answer is appropriate. (Similar considerations are calculated to establish that the property of *breaking a promise* or of *not keeping a promise* and the properties of *discharging* or *failing to discharge* and *violating* (other) *duties of special obligation* are also consequential properties.)

If we can say that in such cases the *signal's being given* or the *promise's being kept* is a consequential property of some act(s), then the consequentialist will be able to show that many

acts that fulfill duties of special obligation are consequentially evaluable in the way suggested by (vii). The consequentialist position can be put this way: Consider that "chain" of acts linked by "by's" which terminates in the promise's being kept—i.e., that chain of acts that starts "the promise to do such-and-such was kept by doing . . ."; if there is a causal "by" in that chain, one of the consequences of one of the agent's acts is the promise's being kept. And thus the indirect objection, (vii), is satisfied.

Insofar as the consequentialist adopts this move, he is committed to holding that *keeping the promise* or *being an act of promise-keeping* is a morally relevant consequential property. (Similar commitments are required to account for other duties of special obligation in this by-chain way.) It is clear that to hold that such properties are morally relevant is to move away from historical consequentialist positions, and toward historical anti-consequentialist positions. That it is not a consequentialist position is, however, not at all clear.

Even if we countenance this indirect consequentialist rejoinder, it might be objected that the consequentialist has not yet established his case. At most this consequentialist rejoinder shows that *some* acts that fulfill duties of special obligation are consequentially evaluable in the way suggested by (vii). That it does not show that *all* such acts are, in this way, consequentially evaluable can be argued as follows. Consider a case where a person promises to twitch a muscle or think of a cat, and keeps his promise. Is there any room for a causal "by" in the chain of acts that begins "he fulfills his promise to twitch his muscle by . . ." or in the chain of acts starting "he fulfills his promise to think of a cat by . . ." Since the promises are kept simply by twitching a muscle or simply by thinking of a cat, the answer would appear to be No. There is no causal relation here. Strictly similar considerations apply in regard to breaking or not keeping such promises. The strategy used against the indirect consequentialist rejoinder (vii) is to find some act that one can promise to do, therefore, have an obligation to do, can do, and can do *directly*—i.e., not by doing something that causes it, however indirectly, to be done.¹¹

The consequentialist might object that fulfilling the promise to *twitch a muscle* or to *think of a cat* does not fulfill these conditions. I think there are two places for such objections. The first is that these

¹¹ This distinction is at least close to, if not the same as, Arthur Danto's in "What We Can Do," *The Journal of Philosophy*, vol. 60 (1963), pp. 435-445 and "Basic Actions," *American Philosophical Quarterly*, vol. 2 (1965), pp. 141-148.

acts are not done directly. The second is that promises to do such acts cannot be morally binding—because of the moral nature of what must be done to fulfill them. I shall defer this second objection to the next section, (vi) when I examine the direct consequentialist objection.

I do not see any plausible way to argue that such an act is not done directly. Nonetheless, it is important to note that to the extent to which it is reasonable to believe that the following form of determinism is correct, to that extent it is reasonable to believe that such acts are not done directly and also that consequentialism is correct: That form of determinism is that all acts that are not morally indifferent must be consequences of another act or acts of the agent. (Various other deterministic views would support consequentialism, too.)

As intriguing as this view of determinism may be, both in its own right and as a way to shore up (vii), I do not find it plausible. So far then, I believe, the consequentialist has not shown that all acts of promise-keeping are consequentially evaluable by this first part of the indirect rejoinder, (vii).

VI. THE CONSEQUENCES OF DOING WHAT IS RIGHT

In this section I shall examine the consequentialist claim that each morally evaluable act is consequentially evaluable in virtue of *its own* consequences. This is (vi), the direct rejoinder. The consequentialist argument here will be seen to be essentially that of the previous section; but here it will not be assumed that different act descriptions do give different acts. The position advanced here is simply that acts of fulfilling (or failing to fulfill) duties of special obligation have special sorts of consequences in virtue of which alone they are acts of fulfilling (or failing to fulfill) such duties; and thus, in virtue of which they are consequentially evaluable.

Consider the sort of case that is often advanced to show that acts that fulfill duties of special obligation are not consequentially evaluable: Smith owes Jones a debt of reparation, which he fulfills by benefiting Jones; Brown, who owes Jones no such debt, benefits him to the same degree as did Smith. The anti-consequentialist argument draws on such cases by claiming that even though the consequences are exactly the

same in each case—viz., Jones being benefited to the same degree—in only the first case is there an act that discharges a duty; therefore, consequentialism is mistaken.

But are the consequences the same? To be sure the eudaimonic or hedonic consequences are the same. Both Smith's and Brown's acts have the property of *benefiting Jones* to the same degree. But a consequentialist need not be a eudaimonist or hedonist. He can argue that the consequences of the two acts are importantly different: In the first case there is the consequence of *benefit that discharges a duty of reparation*; in the second case, there is only *benefit* (or *benefit that does not discharge a duty of reparation*). The consequentialist would argue that the difference between these two sorts of consequences is just what accounts for the difference in the moral evaluation of these acts. And it is because the first act has its sort of consequence that it is an act that discharges a duty.

As problematic as this sort of claim may be, it works especially well in regard to the duties of reparation and gratitude. For unless an act benefits someone (i.e., produces benefit for him), or satisfies him (i.e., produces satisfaction for him) that act cannot even be an act of discharging either of these duties.¹² But can such a claim be sustained in regard to acts of keeping binding promises? Must such acts have "special" consequences? To answer these questions, I shall catalogue some of those conditions that seem necessary for having made a binding promise.

(1) You have made a binding promise only if you say (or do) the correct thing—e.g., "I promise . . ." (2) Your saying the correct thing creates a binding promise only if you were not forced or deceived into saying that thing. (3) Your freely saying the correct thing creates a binding promise only if what you promised to do is not clearly immoral—or perhaps, only if you or others did not see, or were not unreasonable in not seeing, that it is clearly immoral. (4) Your freely saying the correct thing where what you promise to do is not clearly immoral creates a binding promise only if you had good reason to believe or should have realized that by saying the correct thing you were taken by the relevant people to have committed yourself to do that thing.

The most important necessary condition for the dispute over consequentialism is the one I shall call (5) the *intent condition*. This condition is meant to explain why no binding promise has been made in

¹² Cf. fn. 5, (2).

the following cases even though (1)–(4) are satisfied: knowing (or reasonably believing) that you do not care in the least whether or not I pick up this rock I say to you, satisfying (1)–(4), “I promise you I will pick up this rock”; knowing (or reasonably believing) that you will be annoyed if I push you in the water, I say to you, satisfying (1)–(4), “I promise you I will push you in the water.”

Various intent conditions might be advanced to explain why such promises are not binding. As I shall argue in this section, we can offer one intent condition which is plausible and which, if true, would show consequentialism correct. This intent condition, which thus appears to me to be a sufficient, but not necessary, condition for the truth of consequentialism is: A promise is binding only if the promise (i.e., doing what was promised) was calculated, believed, or intended to benefit the person(s) specified by the promise as beneficiary, or to satisfy this person by satisfying his wants, desires, interests, and so on.¹³ This will be called the *causal intent condition*, for it is not possible, I believe, to benefit or satisfy someone by keeping a promise except by causally producing something—something that might be called benefit or satisfaction.

The above five conditions are necessary, but probably not sufficient, conditions for having made a binding promise. The last condition I wish to mention is not, however, another condition for having made a binding promise; rather it is a condition that acts must satisfy if they are to be acts of keeping binding promises. (Here we might contrast those conditions for owing a debt of gratitude and those conditions for an act's being an act of discharging such a debt.) This is (6) the *spirit condition*: if you cannot carry out the spirit of the promise by doing a given act, then in doing that act, whatever else you are doing, you are not keeping a binding promise; and if you are obligated to do that act, your obligation does not stem from its being an act of keeping that (binding) promise.

The spirit condition is meant to explain cases like the one where I promise you *to practice some music in the next several days*, where it is understood—i.e., where it is part of the spirit of the promise—that I am to do this so that we can play a duet next week. If between the time the promise is made and the time I practice, you break your arm and thus cannot play for several months, I no longer have any obligation at all *to practice in the next several days*.¹⁴ Or consider the case where Jim promises to take Sally to the next dance, but between the making of the promise and the dance they become completely indifferent to each other and Sally thus does not want Jim to take her to the dance. For the same reasons, he would no longer have any obligation to keep that promise.

The spirit condition should be understood against the background of the intent condition. For at least part of the spirit of the promise should include what it was calculated, intended, or believed that the promise would effect. Thus if the causal intent condition is correct, we should consider the following hybrid, the *causal spirit condition*: In doing an act you are keeping a binding promise only if that act produces benefit or satisfaction in the ways specified in the promise—i.e., to the beneficiary and in the ways specified by the promise and in the keeping of the promise: i.e., in the promiser's doing what he promised to do. An act satisfies the causal spirit condition solely in virtue of producing benefit or satisfaction in the ways specified in the promise, and in keeping the promise.

Thus the consequentialist argues that there are certain specifiable consequences that all and only acts of keeping a binding promise must have. And these are, roughly, benefit or satisfaction produced in the ways specified in the promise and produced in keeping the promise. Before arguing that this is plausible, and thus before arguing that the causal intent condition is plausible, it may help to examine a variant of the argument presented in the second paragraph of this section.

¹³ Two comments on this condition. (1) The reason this is not a necessary condition for consequentialism is that consequentialism as such need not say what sort of consequences are morally relevant. All that consequentialism as such requires is that all morally relevant properties for act evaluations be consequential properties. However, it does seem that insofar as we are interested in those properties that are morally relevant for evaluating an act as right, at least some of these properties must not be either connected with harm or simply neutral. (2) As I think will become clear, for this discussion we need not be concerned about exactly who must lack those intentions, beliefs, and so on. Nor need we be concerned about the exact nature of the intended benefit, satisfaction of wants, desires and so on—so long as it is *causally produced*. Nor need we worry about the sort of being or thing that can be the intended beneficiary of the promise. For simplicity, I shall write as if the beneficiary is a person.

¹⁴ Adapted from Pickard-Cambridge, “Two Problems About Duty,” *Mind*, vol. 41 (1932), p. 158. See Ross's discussion of this, *The Foundations of Ethics* (Oxford, 1963), pp. 94ff.

There it was argued by the anti-consequentialist that an act that fulfills a debt of reparation can have exactly the same consequences as an act that does not fulfill such a duty; and thus the consequences of discharging a debt of reparation cannot be the sole ground of such an act's being right or a duty. Here it might be argued that acts that are not acts of keeping a binding promise can causally produce exactly the same sort of benefit or satisfaction—as required by the causal intent condition—as acts by which one keeps binding promises; and thus having such consequences cannot be the sole ground of the rightness of acts of keeping binding promises. In the third paragraph of this section that argument was answered by noting that it required the consequentialist to limit his choice of morally relevant consequences to hedonic or eudaimonic ones; and that the argument was mistaken if we countenance such consequences as *benefit that discharges a debt of reparation*. Here the same sort of reply is given. The consequences which are advanced by the consequentialist as the sole ground of the rightness of acts of keeping binding promises are *not* benefit and satisfaction. Rather, they are roughly, *benefit and satisfaction that are produced in the ways specified by the promise and in keeping the promise*.

I say that the consequences are *roughly* as described since that description allows the consequences of keeping non-binding promises to count here, too. For that description does not require that the promise in question satisfy such conditions as (1)–(5) and so on. But this is easily remedied for we can simply add onto the description of such consequences: *where this promise satisfies conditions (1)–(5) and so on*. (I assume here that we could, in a non-question-begging way, fill out the “and so on.”) And for simplicity I shall use “R-consequences” (for relevant consequences) to stand for those consequences of *benefit and satisfaction that are produced in the ways specified by the promise and in keeping the promise where this promise satisfies conditions (1)–(5) and so on*.

The consequentialist claim now to be examined is thus, (a) all and only acts of keeping binding promises have R-consequences; and (b) it is solely because an act has R-consequences that it is an act of keeping a binding promise.¹⁵ In what follows I shall first discuss (a) and attempt to show that it

is at least plausible; and then I shall attempt to show that (b) is, also, at least plausible.

Philosophers have advanced cases that are calculated to show that acts of keeping binding promises need not benefit or satisfy, or even affect, anyone at all. If these claims are correct, then the causal intent condition is not a necessary condition for having made a binding promise; nor, therefore, is the causal spirit condition a necessary condition for an act's being an act of keeping a binding promise. Such cases, thus, are essential for a resolution of this present dispute over consequentialism. Unfortunately, these cases are, for me at least, more or less moot. Consider this combined death-bed and desert-island case presented by Roderick Firth: Jones and Smith, who are shipwrecked on a desert island, build a raft to effect an escape. While waiting for the seasonal winds to begin, so that they will be blown into shipping lanes, Smith cultivates some flowers. Jones is not interested in such activities, but he is at least somewhat pleased that his friend has a pastime that interests him. A few days before the wind begins, Smith cuts his foot, gets a bad infection, and begins to sink fast. On his death bed, he asks Jones to remain on the island a few days after the winds change to make sure that the flowers germinate successfully. He points out how much they mean to him, and that the favorable winds will last for several months, and finally that if Jones remains to tend them, he will have to stay on the island for only a few days after the wind begins. Jones promises to stay and tend them. Smith dies. End of story.

Of course the question is: Does Jones have any obligation to tend the flowers because of his promise? For an affirmative answer to serve the anti-consequentialist argument, we must assume that Smith will not be benefited or satisfied by whatever Jones does; we must not consider the possible benefit Jones will produce for the flowers (if they can even be beneficiaries of a promise); we must forget about the possible effects on Jones of his keeping or not keeping the promise; we must not consider such a question as whether Jones should, e.g., do penance for giving Smith to believe that he would stay—if Jones never did intend to stay or now thinks of leaving; and so on. Given these various conditions, then, does Jones

¹⁵ At this point we can see how the consequentialist rejoinder of this section is essentially the same as that of Sect. V. If different descriptions do give different acts then R-consequences will be intrinsic to some acts and consequences of others. In the spirit of the argument of Sect. V, it would be claimed that an act that has R-consequences also has as a consequence the act of keeping a binding promise.

have an obligation to remain and tend the flowers because he promised to do so? And if he does have such an obligation, Why? I am unable to answer the first question: I find the case too moot. Thus, I cannot even begin to answer the second. My concept and understanding of promises, and my moral intuitions about promises, are simply not determinate and clear enough to allow me to declare confidently that I believe that Jones does or does not have a *prima facie* obligation to keep that promise.

Contracting may be a more troublesome case for the consequentialist. Suppose that Smith and Jones willingly, knowingly, and so on make the following agreement: Smith will pay Jones a large sum of money and in return Jones will think of a cat at 2:00 P.M. on the next Wednesday. Smith makes it known to Jones that he (Smith) does not want or expect anyone at all to be benefited, satisfied, or even affected by Jones's act. Perhaps Smith shows himself to be eccentric, foolish, or whatever by entering into this contract. But this is not in question. What is in question is whether having accepted the money, Jones has a moral obligation (at least to try) to think of a cat, i.e., whether it would be morally wrong for him not to do so. The question might also be put this way: If Jones knowingly, willingly, and so on accepts the money, but then does not keep his part of the bargain, does he not treat Smith *unfairly*, does he not take *advantage* of him? If Jones does act wrongly and unfairly or if he does take advantage of Smith by so acting, consequentialism would appear mistaken. For there need be no consequential property of Jones's act of *thinking of a cat* or of his *not thinking of a cat* in virtue of which the former is right or obligatory and the latter is wrong.

I find this case more persuasive than the previous one. For we might wish to say that a person has the right to dispose of his goods as he wishes and, at least so long as nothing wrong is done, in exchange for whatever he wishes. On the other hand so long as Smith does not want, expect, require, etc., that he or anyone else be benefited, satisfied, or even affected by Jones's act of thinking of a cat (and Jones knows all this), I am not clear how I would evaluate Jones's not thinking of a cat. I would be very tempted to say that it really did not matter, that it was indifferent. Thus, I think that this case is also more or less moot—but perhaps less moot than the desert island case. (To be sure, had Jones believed that he would

not think of a cat, then in *taking the money* he would act unfairly and wrongly. But this is a different act and a different question.)

Some philosophers would say that in each case Jones has at least some obligation. Insofar as they are correct, then it does seem that not all acts of keeping a binding promise or contract need satisfy the causal spirit condition. But insofar as these and similar cases are moot, so, too, is the claim that acts of keeping binding promises need not satisfy this condition. So far as I can see, clear (i.e., non-moot) cases of morally binding promises and contracts do satisfy the causal intent condition. (Thus, the consequentialist claim of Sect. V that one cannot have a binding promise to think of a cat.)

One way to put this present problem is as a question over the strength of the intent condition. The causal intent condition requires that, e.g., the promiser intend to benefit or satisfy the beneficiary of the promise. If in regard to the above cases we should hold that Jones has some obligation to keep his promise or fulfill his contract then we should reject the causal intent condition. In its place we could have the following intent condition: A promise is binding only if it is intended, calculated, or believed to be to the benefit of the beneficiary of the promise, or *to satisfy his wants, desires, and so on*. The difference between this intent condition and the causal intent condition is that to fulfill the second disjunct of this condition, there need be no causal consequences at all. In the desert island case, if Jones stays and tends the flowers then he will satisfy Smith's wants, desires, or interests—even though so doing will not even affect Smith. And in the contract case, by thinking of a cat Jones will satisfy Smith's desires, wants, or interests—whether or not Smith is thereby affected. And here the "by" is not a causal one, it is a coextensional one.

However, I believe that it is reasonable or plausible to support the causal intent condition rather than this weaker intent condition. For we have, I think, seen that it is at least plausible to hold that all acts of keeping binding promises must have *R*-consequences. That is, we have seen that it is at least plausible to hold that all acts of keeping binding promises must causally produce benefit or satisfaction in the ways promised, by keeping the promise.

Thus, I believe that it has been shown that it is at least plausible to agree with the consequentialist claim that (a) *R*-consequences are necessary and

sufficient for an act's being an act of keeping a binding promise. Another way to put this claim is that having *the property of having R-consequences* is a necessary and sufficient condition for an act's being an act of keeping a binding promise. We should now examine the consequentialist second claim about *R-consequences*, or the property of having *R-consequences*: (b) *R-consequences*, or the property of having *R-consequences*, is that which, and the only thing which, *makes* an act an act of keeping a binding promise.

If we accept (a), as I believe it is plausible to do, the question of whether we should also accept (b) appears to come down to the question of the relative conceptual priority of acts and their properties. For if acts in general, or the acts here relevant, are prior to their properties, or to such properties, (b) will be false; and if properties of acts or these properties are prior to acts or to such acts (b) will be true; and if they are on a par so far as priority is concerned we may not be able to settle this dispute over (b).

As is clear, such disputes are notoriously difficult to settle. And thus it is fortunate that, as will be noted in the last paragraphs of this section, our decision about consequentialism does not rest on settling such a dispute.

But not only may there be a problem about priority; it might be held that it is a mistake to take such properties as having *R-consequences* (or such analogous properties as *being benefit produced by an act of reparation*) to be *bona fide* consequential properties. Perhaps this objection is correct. But here we are faced with another very difficult problem in evaluating the dispute over consequentialism. For it is not clear what is or is not a consequence or a consequential property.

In this regard it may be of some importance to note one feature of *R-consequences* that might lead people to deny that they are real consequences. They are *progenitor-dependent*—or *act-involving*—i.e., the fact that a given consequence is an *R-consequence* logically depends on the fact that it was produced by a certain sort of act. It might be suggested that since *R-consequences* are act-involving, they do not meet those conditions of logical independence from their putative causes that *bona fide* consequences must meet. However, the fact that under the description of "*R-consequences*" they are not logically independent of their causes is not decisive—even if effects must be logically independent of their causes. For we can simply give a new description of them so that

under this description they are logically independent of their causes. We could simply use "consequences of benefit and satisfaction."

It might be countered that under this new description, while the conditions of logical independence are met, these consequences no longer play the role consequentialists require of them. But if they are the same consequences, how could they not play the same role? (What might suggest that they do not is that we cannot recognize their role simply from this new description.)

But there is a way to recur to *R-consequences* that avoids both this present problem and the problem concerning the relative priority of properties and acts. As was pointed out by question (C) of Sect. I, the consequentialist can hold that an act is right solely in virtue of any given *function* of the morally relevant consequences. But now consider a consequentialist who is not concerned to dispute whether *R-consequences* are *bona fide* consequences; but rather is concerned only with consequences of benefit and satisfaction. Again, the consequentialist argues that the promise is kept just in case the benefit or satisfaction is produced in the ways specified by the promise, by the keeping of the promise. But not to acquiesce before the familiar anti-consequentialist argument (presented, e.g., in the second paragraph of this section) this consequentialist holds that the *weight* of consequences of benefit and satisfaction is increased if they are produced by, and are essential to, the keeping of a promise that satisfies conditions (1)–(5) and so on. That is, their weight is increased if they are the benefit or satisfaction component of *R-consequence*.

This consequentialist position is essentially the same as the one using *R-consequences*. But now, the *weightings* of the consequences are, in a way, act-involving. For example, to *know how* to weigh a given consequence we must know how it was produced. It might be a source of legitimate concern if the relevant properties (or our recognition of them as relevant) are logically dependent on the act that is said to have produced them. But why should it be a source of legitimate concern if the way we weigh them is logically dependent on the act that produced them? It is to be noted here that any weighing procedure includes some non-consequential considerations.

It might, finally, be objected that accepting *R-consequences* in *either* form trivializes the dispute over consequentialism. To the extent that it does this to the dispute over consequentialism *as such*

as opposed to that over various consequentialisms or that over the question of which properties are relevant for act evaluations, I welcome this effect of *R*-consequences.

VII. OTHER INDIRECT REJOINDERS

In this section I shall cursorily examine several other indirect consequentialist rejoinders. My justification for treating them so cursorily is that if at points of serious contention the previous consequentialist rejoinders have been successful, then there is really no need to proceed with still other rejoinders; and if the previous ones have been unsuccessful at such points, then it is very unlikely that these will succeed. One cautionary note: All that is required of an indirect rejoinder is that it show how a given act is, e.g., wrong because of the consequences of some other act or acts. Since there are innumerable ways that the consequences of other acts *could* be said to make a given act wrong, it is impossible to canvass all indirect rejoinders. Nonetheless, I believe that the following three are the most plausible of the remaining indirect rejoinders. If they do not exhaust the possibilities of this line, they at least exhaust its probabilities. The three indirect rejoinders that I shall consider are those of *negative consequentialism*, *Mill's consequentialism*, and *rule consequentialism* (on the model of rule utilitarianism).

Negative Consequentialism: How would a consequentialist explain that an act of *not kicking a baby* is right or obligatory? Even if the act of *not kicking a baby* has no consequences it can still be right, thus it cannot be made right by its consequences. Clearly, the consequentialist reply is that *not* doing that act—i.e., not *not kicking the baby*, i.e., kicking the baby—would have consequences that make it wrong. The general claim is that if an act has consequences that make doing it wrong, this in itself makes the act(s) of *not* doing that act right or obligatory.

Even if this response is correct, as it seems to me to be, it will not advance the consequentialist cause very much. At best it will show how various *unproblematic* acts are consequentially evaluable—e.g., the act of *not kicking the baby*. It will not show that the problematic acts advanced by the anti-consequentialist in Sect. V and VI are consequen-

tially evaluable. For those problematic acts—e.g., promise-keeping—are not acts the *non-doing* (or the doing) of which is clearly consequentially evaluable; i.e., it is not clear that the non-doing (or the doing) of these problematic acts need have morally relevant consequences.

Mill's Consequentialism: As I read Mill, he held that an act is obligatory just in case being compelled to do it or being punished for not doing it produces the proper balance of hedonic consequences.¹⁶ For example, he writes:

We do not call anything wrong unless we mean to imply that a person ought to be punished in some way for doing it—if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience.

Although "punishment" may be used here in an over-generous way, Mill's claim seems largely correct. Leaving aside the question of excuses (which can probably be handled consequentially if wrongness can) it does seem that if someone does act wrongly then he does deserve at least one of the above sorts of punishment—if only the reproaches of his own conscience. Many critics of the so-called utilitarian theory of punishment have argued that it is at least possible to get exactly the same sort of consequences by punishing a man who has acted wrongly and a man who has not. However, if the argument of the last section is well-taken, it is not clear that this charge can be sustained. But even if it cannot, it seems to me that Mill will have provided, at best, only a necessary and sufficient condition for having acted wrongly, and not a condition in virtue of which an act is wrong or one acts wrongly. For I would say that the fact that a man has acted wrongly is what makes him deserving of punishment; I would deny that the fact that he is deserving of punishment is what makes his act wrong.

Rule Consequentialism: Finally we must consider the consequentialist appeal to rule consequentialism, on the model of rule utilitarianism. To show rule utilitarianism untenable, one could do no better than simply cite David Lyons' powerful argument found in *Forms and Limits of Utilitarianism*.¹⁷ The fact that rule utilitarianism, in its various forms, is untenable does not show that

¹⁶ *Utilitarianism*, ch. v, paragraphs 12 ff, esp. para. 14, from which the following quotation is drawn. Although it may well be true that "produce" should be prefaced by "tend to," I shall ignore this move toward rule utilitarianism. My sole reason for doing this is that I consider rule consequentialism below, and here I wish to discuss a different feature of Mill's utilitarianism.

¹⁷ David Lyons, *Forms and Limits of Utilitarianism* (Oxford, 1965).

rule consequentialism is also untenable—just as the untenability of, e.g., Sidgwick's utilitarianism does not show consequentialism as such untenable. However, the argument I shall advance below is calculated to show that rule consequentialism is almost certainly untenable if act consequentialism is. (I believe that the following is to be found in Lyons' argument.)

As was suggested, it is at least plausible to believe that there are binding death-bed promises, the fulfilling of which need have no morally relevant consequences. Given this, I shall show that it seems equally plausible to believe that rule consequentialism is mistaken. Consider the general or typical or paradigm act of keeping such death-bed promises. I suggest that there is no reason to believe that the general or typical or paradigm fulfillment of such promises need have consequences, much less morally relevant consequences. (Hereafter for brevity I shall speak only of the *general* fulfillment of such promises, rather than the *general* or *typical* or *paradigm* fulfillment. At most, details of my argument are affected.)

It might be countered that the rule "behind" acts of keeping binding death-bed promises is not cast in terms of death-bed promises, but rather in terms of promises *in general*. That is, it is because the general keeping of promises has morally relevant consequences that death-bed promises are binding.

But now there is the following serious problem for the rule consequentialist. His position is that if the general performance of act tokens of type *T* has a given sort of morally relevant consequence then it is, e.g., wrong to do any token of *T*—any *T*, for short. But as shown by death-bed promises at least some of these *T*'s need not contribute to the production of that morally relevant sort of consequence. But now consider those *T*'s (or that proportion of *T*'s) that do not contribute to the production of those consequences—call them the *effectless T*'s and call the other *T*'s the *effectual T*'s. Consider, now, some other non-empty sort *S* of acts that contains at least some non-*T*'s such that if *T*'s as well as *S*'s are generally done, the same morally relevant consequences result as when only *T*'s are done. Let *R* be the sort of acts composed of *S*'s and *T*'s. By hypothesis, the general doing of *R*'s produces exactly the same morally relevant consequences as the general doing of *T*'s. From this it should follow that just as doing any *T* is wrong, so, too, doing any *R* is wrong. But this cannot be

correct. For the *S*'s might be such morally in different acts as thinking of a cat, or they might be such supposedly morally obligatory acts as keeping a death-bed promise to think of a cat.

The consequentialist cannot reject these criticisms by telling us to consider the "smallest effectual" sort of acts in question—i.e., a sort of act the general performance of which produces morally relevant consequences, where all effectless acts have been removed. This would, for example, preclude holding that in a rule consequential way we can evaluate the keeping of an effectless death-bed promise. In short, to ask us to consider only the smallest effectual sorts of acts is to ask us to collapse rule consequentialism into act consequentialism.

I believe that this (and Lyons') objection to rule consequentialism could be met only by showing how the sort *R* of acts is not the kind of act sort that the rule consequentialist is interested in. But then the rule consequentialist must give us ways to identify those kinds of act sorts he is interested in: ways that prevent us from making act sorts "too small" (e.g., only effectual *T*'s) and ways that prevent us from making them "too large" (e.g., all *T*'s and *S*'s). In its broadest and most interesting form, doing this would require an exposition of a theory of natural kinds of acts, of natural act types. Until this is done, we have no reason to believe that rule consequentialism succeeds where or if act consequentialism fails.

VIII. CONCLUDING COMMENTS

This paper has been concerned with showing that the dispute over consequentialism cannot be clearly settled. For, involved in that dispute are the difficult, ultimately metaphysical, problems surrounding the nature of an act, of an intrinsic property, and of a consequential property. Also involved in that dispute are the difficult moral questions surrounding the nature of a binding promise. This paper has shown how these various problems figure in the dispute over consequentialism and has also indicated some ways in which they may be handled to begin to resolve the dispute over consequentialism. Until these problems are solved in a far more definite way, it is impossible to declare with confidence either for or against consequentialism.

It is my view, however, that the crucial problem for this dispute is the one surrounding the nature of an act of keeping a binding promise. If benefit

or satisfaction must be a consequence of such an act (or must be thought to be a consequence of such an act), then it may well be that some form of consequentialism is correct—e.g., the consequentialism of Sect. V (bolstered with the negative consequentialism of Sect. VII) or the consequentialism of Sect. VI (similarly bolstered), or some combination of these. If acts of keeping binding promises need not have (or be thought to have) benefit or satisfaction as a consequence,

then it seems very unlikely that any form of consequentialism is correct. But I think that this question about binding promises is more or less moot. At the least, I have seen no conclusive considerations on either side of the question. Thus it seems to me that the strongest justifiable conclusion concerning the dispute over consequentialism is that it may be somewhat more reasonable to believe that there is no correct form of consequentialism than it is to believe that there is one.¹⁸

Cornell University

Received October 28, 1968

¹⁸ I am happy to acknowledge the very great help—far beyond the call of duty or friendship—David Lyons has given me in preparing this paper. I also wish to thank Norman Kretzmann for helping to make this paper less unreadable than it was.

III. TENSED PROPOSITIONS AS PREDICATES

ARTHUR N. PRIOR

PRIMA FACIE, we may divide propositions, or ostensible propositions, into two sorts. There are those, such as "Socrates is sitting down" or "Socrates was sitting down," of which it obviously makes sense to ask "When are they true?", though the answer may in some cases be "Always" or "Never." And there are, on the other hand, those of which it does not so obviously make sense to ask this question; one sub-species of these would be exemplified by "Two and two are four," and another by "The date of the Battle of Hastings is 1066." Rescher has suggested that propositions in the first major division (those of which at least some are true at some times and false at others) be called "chronologically indefinite," while those in the second major division he calls "chronologically definite."¹ In what follows, I shall call propositions of the first sort "tensed," and propositions of the second sort "untensed." One view of the relation between the two sorts of propositions is that "untensed" propositions are improperly so called; they are simply tensed propositions which are so obviously either always or never true that to ask *when* they are true is socially (but not logically) inept. The most plausible alternative to this is to say that tensed propositions are not genuine propositions but predicates of instants ("Socrates is sitting down" really means "Socrates is sitting down at —"), the instants which form their subjects being unmentioned, but usually understood to be the time of utterance.

Of these two views about the relations between tensed and tenseless propositions, I am myself a determined adherent of the first, i.e., the view that so-called tenseless propositions are simply a sub-class of those tensed propositions which are always or never true. It is not my present purpose, however, to defend this view, but rather to see how far we can take the opposite view, i.e., the view that tensed propositions are not propositions at all, but predicates of instants. I shall take my departure from a paper of 1967 on "Tense Logic and the Logic of Earlier and Later,"² hereinafter

called TEL, in which I outline a series of formal systems concerned with the notion of a proposition or quasi-proposition p being true at an instant a . In the system which I call System I, and also in a slight enlargement of it which is given no number in TEL but which I shall here call System I⁺, tensed propositions are a fairly restricted class; in particular, propositions to the effect that such and such a tensed proposition is true at such and such a time are not themselves classed, in this system, as tensed propositions. Nor are propositions to the effect that such and such an instant is earlier than such and such another instant. These restrictions are removed in the system which I call System II. And in System III, instants themselves are treated as tensed propositions of a sort (each instant is identified with a tensed proposition which would ordinarily be said to be true at that instant only). Using Systems I and I⁺, in which tensed propositions are syntactically separated from propositions about instants, it is plausible to regard these tensed propositions as predicates of the instants at which they are said to be true; but this way of looking at the matter is hardly consonant with the syntax of System II, and still less with that of System III. I therefore describe these systems, in TEL, as embodying increasing "grades of tense-logical involvement."

In the present paper I make both formal and philosophical revisions of TEL. In the first place, I make the syntax of the various systems more explicit than it is in TEL. Secondly I reaxiomatize them in a way which I hope will be more illuminating (and which will at two points be more correct) than the axiomatizations of TEL. Finally, I show that the view of tensed propositions as predicates, though I still think it misguided, can be carried further than is admitted in TEL. I do this last by making certain modifications to the more "tense-logically involved" of my systems, and tracing some of their consequences. In the course of these maneuvers, I hope that some new light may be thrown not only on tense logic but also on predicate calculus.

¹ N. Rescher, "On the Logic of Chronological Propositions," *Mind*, vol. 75 (1966), pp. 75-96.

² A. N. Prior, *Papers on Time and Tense* (Oxford, 1968), ch. XI.

I. THE SYSTEMS I, I⁺, II, IIIa, IIIb, IIIc⁺
AND IIIc

The vocabulary of System I consists of the instant variables a, b, c , etc., the tensed propositional variables p, q, r , etc., and the undefined constants C, N, G, H, I, U, T , and Π . Tensed propositions are defined as follows:

- (i) The variables p, q, r , etc., are tensed propositions.
- (ii) If ϕ and ψ are tensed propositions, so are $C\phi\psi, N\phi, G\phi$, and $H\phi$.

Informally, $C\phi\psi$ is to be read as "If ϕ then ψ ," $N\phi$ as "Not ψ ," $G\phi$ as "It will always be that ϕ " and $H\phi$ as "It has always been that ϕ ." Other truth-functions, in particular $A\phi\psi$ ("Either ϕ or ψ "), $K\phi\psi$ ("Both ϕ and ψ ") and $E\phi\psi$ ("If and only if ϕ then ψ "), may be defined in the usual ways, and $F\phi$ (for "It will be that ϕ ") and $P\phi$ (for "It has been that ϕ ") as $NGN\phi$ and $NHN\phi$ respectively. Untensed propositions are defined as follows:

- (i) If u and v are instant variables, Iuv and Uuv are untensed propositions.
- (ii) If u is an instant variable and ϕ a tensed proposition, $Tu\phi$ is an untensed proposition.
- (iii) If α and β are untensed propositions and u is an instant variable, $C\alpha\beta, Na$ and $\Pi u\alpha$ are untensed propositions.

Informally, Iuv is to be read as " u is the same instant as v ," Uuv as " u is earlier than v ," $Tu\phi$ as "It is the case at u that ϕ ," and $\Pi u\alpha$ as "For all u, α ." C and N are for "If" and "Not," and the other truth-functions definable in terms of them, as before; and $\Sigma u\alpha$, for "For some u, α ," is definable as $N\Pi uNa$.

All the theses of the system are untensed propositions, and the postulates consist of the following axioms subjoined to ordinary propositional calculus, quantification theory, and identity theory:

- ET1. $ETaCpqCTapTaq$
- ET2. $ETaNpNTap$
- UT1. $ETaGp\Pi bCUabTbp$
- UT2. $ETaHp\Pi bCUbaTbp$

Or in words:

- ET1. It is true at a that if p then q , if and only if, if it is true at a that p it is true at a that q .
- ET2. It is true at a that not p , if and only if it is not true at a that p .

³ See, e.g., A. N. Prior, *ibid.*, p. 116.

UT1. It is true at a that it will always be the case that p , if and only if it is true at all instants later than a that p .

UT2. It is true at a that it has always been the case that p , if and only if it is true at all instants earlier than a that p .

As observed in TEL, the symbol T is superfluous in this system; we could write $Tu\phi$ as ϕu and regard the axiom as defining complex predicates in terms of propositional complications, thus:

- ET1. $(Cpq)a = Cpaqa$ Df.
- ET2. $(Np)a = N(pa)$ Df.
- UT1. $(Gp)a = \Pi bCUabpb$ Df.
- UT2. $(Hp)a = \Pi bCUbapb$ Df.

Or in words:

- ET1. " a is an if- p -then- q -ish instant" = "If a is p -ish then a is q -ish."
- ET2. " a is a not- p -ish instant" = " a is not a p -ish instant".
- ET3. " a is a p -for-evermore-ish instant" = "All instants later than a are p -ish".
- ET4. " a is a p -through-all-the-past-ish instant" = "All instants earlier than a are p -ish".

It is also stated, erroneously, in TEL that the equivalence ET1 can be weakened to an implication; in fact this weakening is only possible in System II. The proof of ET1 from the corresponding implication requires the rule that if a is a thesis so is Taa ; in System I either the a or the Taa of this rule will not be an untensed proposition as above defined. However, for System I as now axiomatized we can prove the metatheorem that if the tensed proposition ϕ is tautological in form, $Ta\phi$ is a thesis. (Just use ET1 and ET2 to equate the tensed tautology, preceded by Ta , to an untensed one; e.g., $TaCp\phi$ to $CTapT\phi$). And, as mentioned in TEL and elsewhere, the addition of UT1 and 2 enables us to prove, preceded by Ta , all theses of Lemmon's minimal tense-logic K_t , the postulates of which I need not repeat here.³

In System I⁺, we enlarge the class of tensed propositions by the following additions:

- (iii) If one of ϕ and ψ is a tensed proposition and the other untensed, $C\phi\psi$ is a tensed proposition.
- (iv) If u is an instant variable and ϕ a tensed proposition, $\Pi u\phi$ is a tensed proposition.

And for these new tensed propositions, we add to the axioms of System I the following axiom-schemata:

ET1.1. $ETaCapCaTap$
 ET1.2. $ETaCpaCTapa$
 ET1.3. $ETaIbb\phi IbbTap$

where ϕ is a tensed and a an untensed proposition. (The necessity for ET1.3 was overlooked in TEL). These equivalences can also be replaced by definitions, and we can regard this enlargement as simply allowing us to form complex predicates not only from other predicates but from a combination of these with propositions. (This is normal procedure in predicate calculus). And it may be noted that each of the schemata ET1.1-3 equates something which is ill-formed in System I with something which is well-formed there; thus $TaCIbbTbb$ (e.g., "It is true at a that if Socrates is at all times sitting down then Socrates is sitting down") is equated by ET1.3 with $CIbbTbbTap$ ("If it is true at all times that Socrates is sitting down then it is true at a that he is").

In System II the entire class of untensed propositions is absorbed into the class of tensed ones, and tensed propositions are admitted as theses. This makes the syntactical enlargements of System I⁺ redundant, and its postulates ET1.1-3 are also made redundant by the postulates which System II adds to System I, namely

RET: $\vdash a$ if and only if $\vdash Taa$, for a not free in a .
 ET3. $ETaTbbTbb$
 ET4. $ETaIbbTbbIbbTbb$
 ET5. $ETaUbcUbc$.

These postulates are equivalent to those given for System II in TEL, and, as there observed, the ET portion of the system is equivalent to a system of Rescher's. We cannot now dispense with the functor T in favor of simple predication without using propositions as predicates, i.e., having forms like $(pa)b$ and $((pa)b)c$. However, each of the axioms above equates a formula which is ill-formed in System I with one which is well-formed there, and the rule has a similar function.

Finally, in System III, the class of tensed propositions absorbs even instants, and for instants considered as propositions we have the one further axiom (which makes ET5 redundant).

ET6. $ETabIab$,

asserting that the "instant" b is "true at" a if and only if it is identical with a . The identity function is of course taken to have its usual postulates (say Iaa and the schema $CIabCa\beta$, where β differs from a only in having a b in place of an a). In TEL, I is not thus used as a primitive, though the possi-

bility of doing so is mentioned. System III seems as far removed as possible from the representation of tensed propositions as predicates of instants—instants as objects have just disappeared. But (1) ET6 is like earlier extensions of System I in equating a formula which is ill-formed in that system with one which is well-formed there, though the I of Iab must be re-interpreted as some sort of connective when a and b are read as propositions; and (2) a very slight modification to System III will make it much easier to interpret its tensed propositions as predicates.

The modification I have in mind consists in dropping the identification of an instant with a tensed proposition true at that instant only, and introducing a functor S which forms that proposition from that instant. The tensed proposition Sa may be read as "It is a ," in the sense of "it is" in which we say "It is 5 o'clock," "It is Thursday," "It is the 5th of June," etc. And instead of adding ET6 to System II, we add

S1. $ETaSbIab$,

"That it is b is the case at a if and only if a and b are the same instant." Moreover, the form Sa and the axiom S1 could be added, not to System II, but to System I⁺ or even System I without doing violence to the syntax of those systems; though if the form Sa is introduced into System I as a tensed proposition, it is natural to introduce quantifications of tensed propositions, and with them ET1.3, into that system.

For these various possible new systems I propose to adopt the following names:

- (1) "System IIIa" for System III as above presented, i.e., System II with instants counted as tensed propositions and ET6 added to the postulates.
- (2) "System IIIb" for System II with the following added to the definition of a tensed proposition: "If u is an instant-variable, Su is a tensed proposition"; and with S1 added to the postulates.
- (3) "System IIIc⁺" for System I⁺ with the same additions to the definition of "tensed proposition" and to the postulates.
- (4) "System IIIc" for System I with the following added to the definition of a tensed proposition: "If u is an instant-variable and ϕ a tensed proposition, Su and $Iu\phi$ are tensed propositions"; and with ET1.3 and S1 added to the postulates.

In Systems IIIc⁺ and IIIc, it may be noted, as in

Systems I^+ and I , the symbol T can be dispensed with in favor of predication, and S_1 replaced by the definition

$$(Sb)a = Iab. \text{ Df.}$$

II. QUASI-MODAL FRAGMENTS OF THE SYSTEMS

Where ϕ does not contain free a , we may abridge $\Pi a T a \phi$ ("It is true at all times that ϕ ") to $L\phi$, and it is pointed out in TEL that for an L thus defined we may prove in System II the Lewis modal system S_5 , and also $CLpGp$ and $CLpHp$. What is not investigated in TEL is the result of introducing this definition of L into Systems I^+ and I . The most important point to note about this is that in these systems L is a "heterogeneous" functor, forming untensed propositions from tensed ones. $CLCpqCLpLq$, $CLCpqCLCqrLCpr$, and (with M for NLN) $CLCpqCMq$, $CKLpMqMKpq$ and $CLpMp$ are examples of formulae which remain well-formed, untensed, and theses in these systems; but $CLpp$, $CLpGp$, and $CLpHp$ are ill-formed in System I (being implications with untensed antecedents and tensed consequents), while in System I^+ they are tensed and therefore not theses. However, $LCLpp$, $LCLpGp$, and $LCLpHp$ are theses of System I^+ , and $CLpLGp$ and $CLpLHp$ even of System I .

The fragments of Systems I and I^+ which use only C , N , L , and tensed propositional variables may be regarded as notational variants of certain forms of predicate calculus without name-variables developed by G. E. Hughes and D. G. Longey. In their published textbook,⁴ Hughes and Longey give a calculus in which predicates consist of predicate-letters f , g , h , etc., and truth-functional complexes of these, atomic propositions are formed by prefixing U or E (in effect universal and existential quantifiers) to these, and molecular propositions in the usual truth-functional ways. As postulates we have all tautologies formulable in the language, all tautological truth-functional complexes of predicates with the whole preceded by U , and the following axioms (for use with predicate-substitution and detachment):

$$A1. U(f \supset g) \supset (Uf \supset Ug)$$

$$A2. Uf \supset Ef$$

Without A_2 (which expresses the non-emptiness of the universe), this system would appear to be equivalent to an earlier one with the same syntax, namely von Wright's "Quantified Logic of Properties."⁵ Von Wright also has the idea of classifying the operators U and E as "modalities" ("existential" modalities).⁶ At all events, a mere re-writing of the Hughes-Longey postulates in our symbolism will give a basis for the $C-N-L-p$ portion of System I , namely: All tautologies formulable in the language, all tautological complexes of tensed propositions with the whole preceded by L , and the axioms $CLCpqCLpLq$ and $CLpMp$.

Hughes has also axiomatized⁷ an extension of the above predicate calculus in which propositional as well as predicate variables are used in the formation of complex predicates (an implication of which one component is a predicate counts as a predicate). The axioms are all tautologies, all tautologously formed predicates (including those with propositional components) preceded by U , and the axioms

$$A1.1. U(p \supset f) \supset (p \supset Uf)$$

$$A1.2. U(f \supset g) \supset U(Uf \supset g)$$

$$A2. Uf \supset Ef.$$

This gives for the $C-N-L-p$ fragment of System I^+ : all tautologies, all tautologously formed tensed propositions preceded as a whole by L , the axiom schema $CLCpCaLp$ (where a is untensed) and the axioms $CLCpqLCLp$ and $CLpMp$. ($LCLpp$ is obtained from $LCpp$ and the first axiom, and $CLCpqCLpLq$ from the first axiom and the schema).

Nested modalities like LLp and LMp are of course ill-formed in both these systems, but they are not simply modal logics with such nestings ruled out (of the sort studied, e.g., by J. L. Pollock,⁸ and earlier but more sketchily by myself),⁹ since even $CLpp$ is not propositional enough to figure as a thesis.

If we add the tense-operators G and H , the resulting enlarged fragments of both systems will contain all theses of the tense-logic K_t preceded as a whole by L , and may be axiomatized by adding to the above postulates, besides these L -preceded K_t theses, in System I $CLpLGp$ and $CLpLHp$, and

⁴ G. E. Hughes and D. G. Longey, *Elements of Formal Logic* (New York, 1965).

⁵ G. H. von Wright, *Logical Studies* (London, 1957), second paper ("On the Idea of Logical Truth(I)").

⁶ G. H. von Wright, *An Essay in Modal Logic* (Amsterdam, 1951), ch. II.

⁷ See A. N. Prior, critical notice of Hughes and Longey's *Elements of Formal Logic*, *Australasian Journal of Philosophy*, vol. 44 (1966), pp. 224-231.

⁸ John L. Pollock, "Basic Modal Logic," *The Journal of Symbolic Logic*, vol. 32 (1967), pp. 355-365.

⁹ A. N. Prior, *Time and Modality* (Oxford, 1957), Appendix C.

in System I+ $LCLpGp$ and $LCLpHp$ (from which $CLpLGp$ and $CLpLHp$ are obtainable by the schema $CLCapCaLp$, with Lp for a and Gp and Hp for p). One way of obtaining the K_i theses preceded by L , in either system, would be to lay down a sufficient set of K_i axioms each preceded by L , and then derive the theorems (a) by using substitution, $CLCpqCLpLq$ and detachment, and (b) where K_i theses are obtained by using the rules to infer $\vdash G\phi$ and $\vdash H\phi$ from $\vdash \phi$, to obtain the corresponding L -preceded theses by $CLpLGp$ and $CLpLHp$. (Tautological tensed formulae preceded by L could be similarly obtained by laying down L -preceded tensed propositional-calculus axioms and using substitution, $CLCpqCLpLq$ and detachment).

There is not much advantage in using the richer syntax of System I+, but there is a little. Suppose, for example, we wish to add to either of the systems the postulate that there is a unique time-series, and that a linear one. Using instant variables, we can express this in the full vocabularies of both systems by $AAUabUbaIab$, asserting of any instants a and b that a is either earlier than or later than or identical with b ; or by $CTaKKpHpGpLp$, asserting that if it is true at any instant a that it is and always has been and always will be the case that p , then p is true at all times. But in the vocabulary of System I+ we can express this assumption in pure $L-C-N-G-H-p$, by $LCKKpHpGpLp$.

In both systems, G and H are "extensional" functions of their arguments, in the sense that we have $CLEpqLEGpGq$ and $CLEpqLEHpHq$. If we read these arguments as predicates, this means that G and H are extensional functions in the standard sense, since these laws just mean $CIIaEpaqaIIaE(Gp)a(Gq)a$ and $CIIaEpaqaIIaE(Hp)a(Hq)a$.

Considered as bits of predicate calculus without name-variables, the extensions with G and H are of some interest. The Hughes-Londes pedagogical program of delaying the introduction of name-variables as long as possible begins to run into difficulties when we move from one-place to two-place predicates (how, e.g., without using name-variables, do we distinguish between $IIa\SbUab$ and $IIb\SaUab$?), and in fact they make no attempt to carry it that far; but here *some* of the logic of the two-place predicate U is brought in via a pair of operators forming one-place predicates from other one-place predicates—in particular, forming

"—has only p -ish instants following" and "—has only p -ish instants preceding" from the predicate-of-instants "—is p -ish." The procedure may be generalized by replacing any relation R by operators, on a predicate f , forming "—is R only to what is f ," and "—is R 'd only by what is f ."¹⁰ The $L-C-N-G-H-p$ logics we have been considering give a general calculus for such pairs of operators, and some well-known extensions of this logic enable us to state, within this type of calculus, certain properties of the "hidden" relation. For example, it is known that if we add to System I the postulate that U is transitive, we can prove not just the theses of K_i preceded by Ta but, preceded by Ta , the theses of $K_i + CGpGGp$; and in the present fragment of System I, the transitivity of the unmentioned U may be expressed by adding the axiom $LCGpGGp$. And the replacement of U by G and H can be taken still further in the Systems III.

III. THE TENSE-LOGICAL CONSTRUCTION OF THE EARLIER-LATER LOGIC

In what I am now calling System IIIa, it was shown in TEL that we may prove $ETa\phi LCa\phi$, and that the function LE satisfies the normal postulates for identity within the system. This suggests a re-axiomatization with L as primitive and T and I defined as LC and LE respectively. With L , C , and N primitive, and the definitions just given, we obtain a system equivalent to the $T-I$ portion of IIIa if we add to S5 for L three special postulates for "instant-propositions," namely Ma , $ALCa\phi LCa\phi Np$ and Σaa .

Further, in System IIIa, with UT1 and 2 included, we can prove $EUabTaFb$ and $EUabTbPa$ (a is earlier than b if and only if it is true at a that it will be that b , and if and only if it is true at b that it has been that a). This suggests dropping U as a primitive and defining Uab either as $TaFb$ ($LCaFb$) or as $TbPa$ ($LCbPa$). If we do this, and add K_i for G and H , with the mixing postulates $CLpGp$ and $CLpHp$, to the $L-C-N$ postulates just mentioned, or to the $T-I$ postulates to which they are equivalent, UT1 and 2 become provable. Thus the elementary logic of the earlier-later relation U can be developed from the logic of the tenses G and H , given the above postulates for L or for T and I . Moreover, those conditions on U which have "reflections" in tense-logic can be proved from those reflections; and even those which have

¹⁰ For beginnings of such generalizations, see A. N. Prior, *Papers on Time and Tense*, op. cit., ch. XII, and "Egocentric Logic," *Noûs*, vol. 2 (1968), pp. 191-207.

none can be reformulated in terms of G , H , and L .

These maneuvers, however, presuppose both the identification of instants with instantaneous tensed propositions (the peculiarity of System IIIa) and the identification of T , U , and I functions with omnitemporal tensed propositions (a feature of System II also). The question which we may now ask is: How far can these maneuvers be reproduced in systems such as IIIc⁺ and c, which do not have these peculiarities, and in which tensed propositions are in effect treated as predicates of instants (just as they are in I⁺ and I)?

The fact that the postulates of the richer systems can be given as simply equating new forms with old is an encouraging sign for this enterprise, and it turns out that practically everything required can be carried out even in the least "tense-logically involved" System IIIc. We can prove, even in System IIIc, $ETapLCSap$, $EUabTaFSb$, $EUabTbPSa$, MSa , $ALCSapLCSaNP$, and $LEaSa$ (the plain ΣaSa , being tensed, is not suitable for a thesis in System IIIc). We can also prove, even in IIIc, the "Barcan" schema $EIIbL\phi LIIb\phi$; an important point because in System II (on which Systems IIIa and IIIb are based) this schema can be represented as a by-product of quantification theory and S5 for L , but it is not provable from quantification theory and the L - C - N - p fragment of System I (on which IIIc is based).

We can now rebuild IIIc in various ways. In the least radical way, we may dispense with T in favor of predication, i.e., replace Tap by pa , and (1) introduce into predicate calculus with identity the definitions $(Cpq)a = Cpaqa$, $(Np)a = Npa$, $(Sb)a = Iab$; and then (2) instead of adding a primitive U (with no axioms) and defining $(Gp)a$ and $(Hp)a$ as $IIbCUabpb$ and $IIbCUbapb$, we could define Uab as $(FSb)a$ or $(PSa)b$, and for the undefined G and H lay down the axioms for K , each predicated of an arbitrary instant a , and the further axioms $CIIbpb(Gp)a$ and $CIIpTb(Hb)a$ (cf. $CLpGp$ and $CLpHp$ in System II and its extensions); equivalences corresponding to the definitions of G and H are then provable. We gain very little, of course, by this; the U -primitive version of System IIIc is obviously more compact; the re-shuffling merely helps to show the sufficiency of K , for a tense-logic presupposing no special conditions on U .

It is possible, however, to proceed quite differently, and more interestingly. We may present the System IIIc as indeed a disguised predicate calculus, but of an extended Hughes-Londesort;

that is, if we may so speak, a predicate calculus without predication. We shall have, indeed, instant variables and quantifiers binding them (our system could not be equated with IIIc without them), but these instant-variables will not figure as subjects of predication—we shall not represent Tap as a variant of pa ; but differently. Instant-variables will occur only as constituents of predicates (or tensed propositions) of the form Su , and Tap will be defined as $LCSap$, i.e., in effect, as the formal implication of the predicate p by the predicate Sa . Nor will S be defined by equating $(Sb)a$ with Iab ; on the contrary, S will be one of our primitives and Iab defined as $LESa Sb$, i.e., as the formal equivalence of the uniquely instantiated predicates Sa and Sb . L will be primitive, and the only undefined way of constructing a genuine or untensed proposition from a predicate, i.e., tensed proposition, will be by prefixing L to it.

We now build up the postulates of the system in the following steps:

(1) We axiomatize the L - C - N - p fragment of System I as in the last section.

(2) We introduce instant-variables, quantifiers binding them, and S , with the three axioms MSa , $ALCSapLCSaNP$, and $LEaSa$; the rules

$$II1: Ca\beta \rightarrow CIIua\beta$$

$$II2: Ca\beta \rightarrow CaIIu\beta, \text{ for } u \text{ not free in } \alpha$$

$$LII1: LC\phi\psi \rightarrow LCIIu\phi\psi$$

$$LII2: LC\phi\psi \rightarrow LC\phi IIu\psi, \text{ for } u \text{ not free in } \phi,$$

(where α , β are untensed propositions and ϕ , ψ tensed); and the schema $CIIuL\phi LIIu\phi$ (the converse being provable).

(3) We introduce the predicate-formers (tensed-proposition-formers) G and H with postulates as in the last section, i.e., L -preceded K , $CLpLGp$ and $CLpLHp$.

(4) We define Tap as $LCSap$, Iab as $LESa Sb$, and Uab as $LCSaFSb$ ($TaFSb$).

All these postulates (including equivalences corresponding to the definitions) are provable in our original version of IIIc; and the latter (including Iaa and the schema $CIabCa\beta$, where a and β are untensed propositions of the system and β differs from a only in a replacement of a by b) is provable from them.

If we construct a similar version of System IIIc⁺ (with (1) and (3) adjusted as in the last section, and with it understood that in $LII1$ and $LII2$ at least one and possibly but not necessarily both of ϕ and ψ must be tensed), we can omit the Barcan schema as provable, though we still need $LEaSa$.

The predicates *Su* function in these systems like Quine's "Pegasizes." Their unique instantiation is guaranteed by their special axioms. *MSa* asserts in effect that at least one individual (instant) has the property *Sa*, and *ALCSapLCSa.Np* that for any property *p* either whatever has *Sa* has *p* or whatever has *Sa* lacks *p*, so that we never have one thing with *Sa* having *p* and another lacking it; i.e., at most one thing has *Sa*.

We could go further than this. To symbolize a uniquely occurring predicate (tensed proposition) as *Sa* still suggests something more like "is Pegasus" than "Pegasizes" (our first reading of this form was, in fact, something like "is 5 o'clock," or more precisely "is 5 o'clock G.M.T. on January 14, 1957"), and we could drop this suggestion of an internal structure by not using the symbol *S* at all and treating the *u*'s (unprefixed) as themselves the uniquely occurring predicates (their unique instantiation being, again, secured by the special axioms). This would give us a system, which we might call IIIId, in which the "tensed propositions" of System I are enlarged by treating instants as such propositions, but are enlarged *only* in this way, genuine or untensed propositions (such as *Tab*, *Uab*, *Lp*, and truth-functions and quantifications of these) being still syntactically segregated from tensed ones, and being alone capable of figuring as theses.

IV. FURTHER CONSIDERATIONS REGARDING SYSTEM IIIId

It is tempting to say that, as far as the removal of dubious presuppositions is concerned, in this last system, IIIId, we contrive to get the best of both worlds. Tensed propositions are still resolutely treated as predicates rather than as genuine propositions (so that our "tense-logical involvement" is still minimal); yet instants do not figure in the systems as named entities, so that we do not seem committed to their "existence" in any serious sense. We seem, in fact, to have come up with a type of calculus which might very well suit people who say such things as that logic should detach itself from a thing-quality or substance-accident metaphysic, or that things should be regarded as logical constructions out of events rather than *vice versa*. Against this it might be argued that although Hughes-Londey types of predicate calculus do not mention individuals they presuppose them, and that if tensed propositions in particular are really predicates we cannot avoid, though we may

symbolically glose over, the question as to what they are predicates of. It may also be remarked that even in System IIIId we still *quantify* over instant variables, and on some views this would suggest we are not taking their treatment as predicates very seriously.

My own view on these points is that (a) a thing-quality metaphysics is fine, but that (b) instants are not things, and (c) quantifying over instant-variables does not commit us to the view that instants *are* things. But I do not wish to argue for these positions here, but rather to note that formally System IIIId, as well as System IIIC, *could* be interpreted by taking the thinghood of instants very seriously indeed, and reading its tensed propositions simply as *class* symbols, denoting *sets* of instants, namely the sets of instants "at" which we would normally say that each proposition is true. This interpretation is possible because of the extensionality of the functions *G* and *H*, considered as functions of predicates. It will be simplest to begin by developing this interpretation of System IIIC.

On this reading $N\phi$ obviously denotes the Boolean complement of the set ϕ , and $K\phi\psi$ the intersection of the sets ϕ and ψ . $L\phi$, which is "genuinely propositional" in Systems IIIC and d, does not denote a set but is a sentence asserting that the set ϕ (i.e., the set of instants at which ϕ) is the universal set (the totality of instants); or in brief, that $\phi=1$. The set $C\phi\psi$, i.e., $NK\phi N\psi$, is the complement of the intersection of ϕ with the complement of ψ ; i.e., that part of the totality of instants which lies outside that part of the set ϕ which does not overlap the set ψ , i.e., the union of the set ψ and the complement of ϕ ($NK\phi N\psi = AN\phi\psi$). $LC\phi\psi$ then asserts that this set $NK\phi N\psi=1$, i.e., that $K\phi N\psi=0$, i.e., that no part of the set ϕ fails to overlap the set ψ , i.e., that ϕ is wholly included in ψ (the obvious set-theoretic version of the statement that all instants at which ϕ are instants at which ψ , or that whenever ϕ is true so is ψ). *G* and *H* are functions from sets of instants to sets of instants, such that $G1=H1=1$, $GK\phi\psi=KG\phi G\psi$ and $HK\phi\psi=KH\phi H\psi$, i.e., the *G*-set of the intersection of ϕ and ψ is the intersection of the *G*-sets of ϕ and of ψ , and similarly for *H*. (These conditions are set-theoretically equivalent to laying down $CL\phi LG\psi$, $CL\phi LH\psi$, $LCGC\phi\psi CG\phi G\psi$, and $LCHC\phi\psi CH\phi H\psi$). Further, relating the two functions, we have $LCNGNH\phi\phi$ and $LCNHNG\phi\phi$. It can then be shown that there is a relation-in-extension *U* between the instants in our sets, for

which $G\phi$ is the set of instants such that all instants U 'd by them are in ϕ , and $H\phi$ is the set of instants such that all instants which are U to them are in ϕ ; or in symbols, $G\phi = a: \Pi bCUab\phi b$ and $H\phi = a: \Pi bCUba\phi b$. For this property will be possessed by the relation between instants a and b which consists in a being in the set $NGNSb$, where Sb is the set of which b is the sole member.

In System IIIId, U is not expressible as a relation between instants, since that system does not contain names for instants. However, it has the symbols a, b, c , etc., denoting the corresponding unit sets, and so has the function Uab which consists in the unit set a being included in the set $NGNb$ ($LCaNGNb$), and for U thus defined we can prove that a unit set a is included in $G\phi$ if and only if $\Pi bCUabLC\phi$ and in $H\phi$ if and only if $\Pi bCUbaLCb\phi$. This, on this interpretation of the system, is precisely what we prove when we derive UT1 and 2 from its postulates for G and H .

Thus interpreted, System IIIId is an "algebra" in the sense of some recent papers by E. J. Lemmon,¹¹ and its proofs of theses about U from theses about G and H are simply versions of completeness proofs using such algebras. An

extensional predicate calculus without name-variables is after all just a version of set-theory without symbols for members or for membership, and that in turn is simply a Boolean algebra, though if members of sets are to be represented in it by their unit classes it must be an "atomic" Boolean algebra, i.e., it must contain elements a such that $a \neq 0$ and for any ϕ , if ϕ is included in a then either $\phi = 0$ or $\phi = a$, and such that if any element $\neq 0$ then some atomic element is included in it.¹²

The set-theoretical proofs will of course go through *whatever* the sets are supposed to be sets of, e.g., if their members are teacups, $G\phi$ is the set of teacups such that all more expensive teacups than these ones are in the set ϕ , and $H\phi$ is the set of teacups such that all *less* expensive teacups are in ϕ . This, to my mind, means that we do not need to take the "instants" of the original interpretation terribly seriously after all; for formal completeness proofs we can happily use System IIIc or d and think of teacups, and for metaphysics we can paraphrase the instants away altogether by thinking of IIIc and d as System IIIa with certain not very illuminating types of well-formed formulae (e.g. $TaTb\phi$) omitted.

Balliol College, Oxford

Received October 30, 1968

¹¹ E. J. Lemmon, "Algebraic Semantics for Modal Logics," *The Journal of Symbolic Logic*, vol. 31 (1966), pp. 46-65, and pp. 191-218.

¹² Cf. Tarski, "On the Foundations of Boolean Algebra," *Logic, Semantics and Metamathematics* (Oxford, 1956), pp. 320-341, esp. 2. Tarski does not have special symbols for atoms, but a definition of atomicity and a postulate for it, as above. This corresponds to variants of our systems in which special instant-variables are dispensed with in favor of the function $Q\phi$, for " ϕ is true at one instant only" (cf. TEL p. 128, and my "Egocentric Logic," *Notas*, vol. 2 (1968), last section).

IV. AQUINAS'S THIRD WAY

THOMAS MAUTNER

AQUINAS'S Third Way is intended as a proof of the existence of an independently-necessary being (i.e., a being which derives its necessity from nothing else). Most discussions of the Third Way are only concerned with the question whether it contains some false, or doubtful, premiss.¹ I shall argue that the Third Way is invalid. It can therefore be rejected without any protracted discussion of the truth-value of this or that premiss.²

I.1 INTRODUCTORY SURVEY

The Third Way can be divided into parts. In the first part, the assumption

Everything is contingent
is claimed to entail

Once there was nothing.

In the second part, the assumption

Once there was nothing
is claimed to entail

Nothing exists now.

From these two entailments it is inferred

If everything is contingent then nothing exists now.

Next, it is affirmed

It is false that nothing exists now
and from these last two statements it is inferred

It is not the case that everything is contingent
or, in other words

Some things are necessary.

In the final part, Aquinas argues that

Everything which is necessary has either dependent necessity or independent necessity.

By an infinite regress argument, similar to those of the First and Second Ways, he infers

It is not the case that everything has dependent necessity

and thus

Something has independent necessity.

The Third Way contains no argument to show that there is exactly one thing which has independent necessity, nor that that thing is God.

I shall not discuss the final part of the Third Way. It stands and falls along with similar infinite regress arguments in the First and Second Ways.

The conclusions from the first two parts follow from their premisses. I shall argue that the second part, despite its *prima facie* obviousness, is not valid as it stands, given certain natural assumptions (I.3). As for the first part, I shall argue (I.4) that it contains a logical error. This error is not trivially eliminable. Therefore the Third Way fails.

I.2 TEXT AND TRANSLATION OF THE TWO FIRST PARTS OF THE THIRD WAY³

- L1 *Invenimus enim in rebus quaedam quae sunt
possibilia esse et non esse,*
L2 *cum quaedam inveniantur generari et corrumpi*
L3 *et per consequens [possibilia] esse et non esse.*
L4 *Impossibile est autem omnia quae sunt, talia
[semper] esse,*
L5 *quia quod possibile est non esse quandoque non est.*

¹ See Sect. II.3.

² This has been argued before (see Sect. II.1). The purpose of this article is to give a faithful account of Aquinas's actual argument rather than a paraphrase of it; to point out another weakness in it which has often been overlooked (Sect. I.3); to expose to criticism a number of prominent recent writers on the topic (Sects. II.2, 3, 4).

³ The Latin text is taken from the Blackfriars' edition of *Summa Theologiae* (London, Eyre and Spottiswoode, 1964). The bracketed lines are variant translations to be discussed below. The brackets in L3 and L4 signify textual variations in some editions of the *Summa*.

Method of reference

The beginning of the Third Way is divided into 12 statements. They will be referred to by parenthesized numbers (1)–(12). Numbers prefixed with an L refer to these statements in Latin.

Numbers prefixed with an E refer to English translations of these statements.

Numbers prefixed with an F refer to formalized versions of these statements.

- L6 *Si igitur omnia sunt possibilia non esse aliquando nihil fuit in rebus.*
 L7 *Sed si hoc est verum etiam nunc nihil esset,*
 L8 *quia quod non est non incipit esse nisi per aliquid quod est.*
 L9 *Si igitur nihil fuit ens, impossibile fuit quod aliquid inciperet esse,*
 L10 *et sic modo nihil esset,*
 L11 *quod patet esse falsum.*
 L12 *Non ergo omnia entia sunt possibilia sed oportet aliquid esse necessarium in rebus.*

- E1 Some of the things we come across can be but need not be,
 E2 for we find them being generated and corrupted,
 E3 and thus sometimes in being and sometimes not.
 (E3₁ and thus in being and not in being.)
 (E3₂ and thus able to be and able not to be.)
 E4 Everything cannot be like this,
 (E4₁ Everything cannot always be like this.)
 E5 for what need not be at some time is not.
 (E5₁ for a thing that need not be once was not.)
 E6 If therefore everything need not be, then at some time nothing was.
 E7 If this were so, then now nothing is,
 E8 because what is not only begins to be by something which is.
 E9 Therefore, if nothing was, it was impossible for anything to begin to be,
 E10 and then nothing would be.
 E11 But this is obviously false.
 E12 Therefore not everything can be but need not be, but something necessarily is.

Some comments on the translation:

(L3) L3 and E3₁ sound self-contradictory. If we accept the "possibilia" of some Latin editions, both E3 and E3₂ are defensible translations.⁴ I have preferred E3, since it seems to me to be more in line with other parts of the argument. Nothing much hangs on this, since (3) is redundant.

(L4) L4 has a "semper" in some Latin editions. But this hardly makes sense, and the same applies to E4₁. For E4₁ denies that everything is sometimes *contingent* and sometimes not; what has to be denied in this context is obviously that everything

is sometimes *in being* and sometimes not. So E4 is clearly preferable.

(L5) E5₁ is the translation offered by T. McDermott in the Blackfriars' edition. It has no support from L5. His translation makes Aquinas sound more coherent in English than the Latin text warrants. At this point there is a gap in the argument and we will discuss it later (I.4). An incorrect translation does not solve any problems.

I.3 THE SECOND PART⁵

What I call the second part of the Third Way is stated in L7-L10.

It is obvious that Aquinas thinks that (8) entails (9). The "*igitur*" ("Therefore") of L9 (E9) confirms this. Also, if this were not so, (9) would be an underived premiss, and (8) completely superfluous.

(8) reads: (because) what is not only begins to be by something which is

(9) reads: (therefore) if nothing was, it was impossible for anything to begin to be.

Consider now the following statement:

Q If there is a time *u*, at which something begins to be, then at any time before *u* something is.

(9) and *Q* are equivalent.

I now want to show that (8) does not entail (9). For this purpose I introduce the following statements R_p ($p=0, 1, 2, 3, \dots$).

- R_0 There is a thing which begins to be at $T+a$.
 R_1 There is a thing which begins to be at $T+a/k$.
 R_2 There is a thing which begins to be at $T+a/k^2$.
 \vdots
 R_n There is a thing which begins to be at $T+a/k^n$.
 \vdots

In these statements R_p , T denotes a time at which nothing is; a and k are numbers, real or rational; $0 < a < k$; if T denotes a time, then, for all p , $T +$

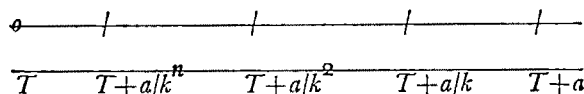
⁴ For instructive discussions of the connection between temporal and modal concepts in Aquinas, see Patterson Brown "St. Thomas' Doctrine of Necessary Being," *The Philosophical Review*, vol. 73 (1964), pp. 76-90, and C. J. F. Williams, "Aristotle and Corruptibility," *Religious Studies*, vol. 1 (1965-66), pp. 95-107, 203-215.

⁵ Particularly with respect to this section, I am indebted to David Bostock of Merton College, Oxford, and George Molnar of the University of Sydney.

a/k^p denotes some time after T . If T , a , k , are constants, then, for different p 's, expressions of the form $T+a/k^p$ will denote different times. Consider now the infinite sequence of the statements R_p . (8) is compatible with their conjunction R . But Q is not: if Q is true then all the R_p 's are false (since at T nothing is, we would have from Q , that nothing begins to be later either). So Q entails that R is false.

Summing up: (8) and R are compatible. (8) and R are self-consistent. Therefore, (8) and R can be true together. Since Q entails that R is false, Q and R cannot be true together. So it is possible that (8) and R are true and Q is false. Q is equivalent to (9). Hence it is possible that (8) and R are true and (9) is false. Therefore it is possible that (8) is true and (9) is false. So (8) does not entail (9).⁶

This can also be illustrated by a diagram:



A '/' through a point on the upper line signifies that something exists at the time marked at the corresponding point on the lower line. An 'o' through a point on the upper line signifies that nothing exists at the time marked at the corresponding point on the lower line. A statement corresponding to (8) would entail that every /-point of the open interval to the right of the o-point on the upper line has a /-point between itself and the o-point. This is so, in the diagram. A statement corresponding to (9) would state that there are no /-points to the right of the o-point. This is not so, in the diagram. It therefore illustrates that a statement corresponding to (8) does not entail a statement corresponding to (9).

(8) and (9) appear to be alternative formulations of the received truth "*ex nihilo nihil*." It has been shown that in fact they are not.

The intuitive plausibility of (8) and (9) seem to me to be equal. Aquinas introduces (8) as an apodeictically true premiss, and derives (9) by a dubious move. If we amend his argument by discarding (8) and introducing (9) as an apodeictically true premiss, this difficulty would be avoided, and the amended argument would not be less plausible than the original argument.

The discussion in this section assumes that time

shares certain features with the set of real or of rational numbers.

I.4 THE FIRST PART: THE MAIN OBJECTION

What I call the first part of the Third Way is stated in L1, L2, L3, L5, and L6.

(5) reads: what need not be at some time is not.

(6) reads: if everything need not be, then at some time nothing was.

In Aquinas' argument, (6) is clearly supposed to follow from (5). But some extra premiss is obviously needed to justify the move from the "is" in (5) to the "was" in (6).

I shall discuss two attempts to supply a missing premiss.

One attempt consists in supplying:

A What need not be at some time was not.

Now we would have, semi-formalized:

F5 $(\forall x)(x \text{ need not be} \supset (\exists t)(x \text{ is not at } t))$

FA $(\forall x)(x \text{ need not be} \supset (\exists t)(x \text{ is not at } t, \text{ and } t \text{ is in the past}))$

F6 $(\forall x)(x \text{ need not be}) \supset (\exists t)(\forall x)(x \text{ is not at } t, \text{ and } t \text{ is in the past})$

This is clearly invalid. What does follow (and gives the illusion of validity) is not F6, but

$(\forall x)(x \text{ need not be}) \supset (\forall x)(\exists t)(x \text{ is not at } t, \text{ and } t \text{ is in the past}).$

Another attempt consists in supplying:

B If at some time nothing is, then at some time nothing was.

Now we could have, semi-formalized:

F5 $(\forall x)(x \text{ need not be} \supset (\exists t)(x \text{ is not at } t))$

FB $(\exists t)(\forall x)(x \text{ is not at } t) \supset (\exists t)(\forall x)(x \text{ is not at } t, \text{ and } t \text{ is in the past})$

F6 $(\forall x)(x \text{ need not be}) \supset (\exists t)(\forall x)(x \text{ is not at } t, \text{ and } t \text{ is in the past})$

This is clearly invalid. It would appear that F6 is obtained from F5 and FB by Hypothetical Syllogism. This is not so. F5 entails

$(\forall x)(x \text{ need not be}) \supset (\forall x)(\exists t)(x \text{ is not at } t)$

but in order to apply Hypothetical Syllogism we need

$(\forall x)(x \text{ need not be}) \supset (\exists t)(\forall x)(x \text{ is not at } t)$

and this is not entailed by F5.

⁶ What (8) does entail is the following statement:

P If there is a time, u , at which something begins to be, then at some time before u something is.

The difference between P and Q helps to explain the fallacious step; the similarity between P and Q helps to explain why it occurs.

These, and all other, attempts to save the argument by justifying the move from "is" to "was" fail, because nothing has been done to justify the invalid switching about of operators from $(\forall x)(\exists t)$ to $(\exists t)(\forall x)$. The argument, therefore, remains invalid. The error⁷ is similar to that of the following examples:

All men are mortal; therefore the human race will become extinct.

and

Every boy loves some girl; therefore there is one girl who is loved by every boy.

Is this a conclusive refutation of the argument under discussion?

If "argument" is used in such a way that an argument, after amendment, is still regarded as the *same* argument as before amendment, then *no* argument can be conclusively refuted, since there are infinitely many possible amendments. I cannot think of any successful amendment of the argument under discussion, and as will be shown in II.2 and II.4, some attempts, the only ones which seem to hold any promise, fail.

If "argument" is used in such a way that an argument, after amendment, is regarded as *different* from what it was before, then an argument can be conclusively refuted, and the preceding discussion amounts to a conclusive refutation.

II.1 SIMILAR VIEWS

Some writers⁸ appear to have been clearly aware of the main objection, and have regarded it as conclusive.⁹ Thus E. A. Burt¹⁰ writes:

One link in this reasoning is felt by many Catholic authors to be insecure, namely, that in which Thomas asserts that if all things without exception were merely contingent, there must have been some time at which nothing existed. They are disposed to admit that even though no one contingent thing exists forever, there might be an unending succession of contingent existences, with no period at which nothing contingent exists. So they prefer to state the argument in a form which does not depend on this link. Their central contention is this: even though it be admitted that there may have been no time at which nothing

existed, if anything is capable of nonexistence it is evidently not self-explanatory; the ground of its existence must lie in something other than itself.

Burt's remark is to the point. But as he puts it, it does not become very clear that this leads to a complete rejection of the Third Way. What "many . . . authors . . . prefer" is not "to state the argument in a form which does not depend on this link," but to state a different argument (one very similar indeed to the argument from contingency in *Summa Contra Gentiles*, I.15.12).

II.2 AN ATTEMPT TO INVALIDATE THE MAIN OBJECTION

R. B. Edwards¹¹ is aware of the main objection but does not regard it as conclusive: Following his own words closely, his argument is:

The cosmological argument is an enthymeme with questionable premisses, but it is not fallacious. St. Thomas Aquinas in his "third proof" argued that the whole of nature at one time did not exist since¹² each of its parts at one time did not exist. This version of the cosmological argument involves no fallacy so long as it is possible to construe it as an enthymeme, as follows:

If all the parts of a whole did not always exist, then the whole itself did not always exist.

All the parts of nature did not always exist; therefore, nature as a whole did not always exist.

There is no 'fallacy' in this argument. I would argue that the *first* premiss is true. I need not go as far as to maintain that it is true because of the meaning of the concepts involved. At least it is true in experience. We certainly cannot find in experience a whole which always existed even though each of its parts did not always exist . . .

I have established that the cosmological argument does not involve any fallacy and that we question the truth (or meaningfulness) of its conclusion only because we question the truth (or meaningfulness) of at least one of its premisses.

This statement is unsatisfactory for two reasons:

(A) Suppose that we have an argument of the form *P* therefore *C*, and that this argument is invalid.—It could *always* be said that *P*, therefore *C*, is valid—although-enthymematic: the suppressed premiss

⁷ Although I have *exposed* the fallacy, I have not *named* it. For it is not clear how fallacies should be labeled, and there is no need to discuss that (wider) problem here.

⁸ E.g., C. J. F. Williams, "God and 'Logical Necessity'," *The Philosophical Quarterly*, vol. 11 (1961), pp. 356-359.

⁹ It could be argued that there is no point in restating what has been said before. It is hoped, however, that Sect. I. 2-3 and II. 2-4 will justify this article.

¹⁰ E. A. Burt, *Types of Religious Philosophy* (rev. ed., New York, Harper & Brothers, 1951), p. 106.

¹¹ R. B. Edwards, "Composition and the Cosmological Argument," *Mind*, vol. 77 (1968), pp. 115-117.

¹² A mistake: instead of "since," read "if."

being Q , where Q entails $P \supset C$. This would be to say that there are no logical fallacies. (At worst, an argument would be enthymematic!)

(B) Correctly, Edwards draws attention to the fact that although in general arguments of the form: "all parts of W are F ; therefore W is F ," are invalid, some arguments of this form may still be valid, not in virtue of having that form, but in virtue of some other feature.—But Edwards has not given any convincing reason why the Third Way should be regarded as a special case. The reason he offers is that the first premiss above (which he takes Aquinas to have tacitly assumed) is true in experience. "Experience does not show us a whole which *always* existed even though each of its parts did not always exist."

Certainly, experience does not show us of *any* whole that it *always* existed. Consequently, statements about such wholes cannot be empirically confirmed or disconfirmed. But surely the absence of disconfirmation cannot be taken as confirmation in a case like this!

As to wholes which exist *permanently* (relatively), experience, on the contrary, *does* show us that their parts need not exist permanently. Take for instance round-the-clock pop-music broadcasts.

Therefore, Edwards has presented no good reason whatever for accepting his supplied premiss.

II.3 WRITERS NEGLECTING THE MAIN OBJECTION

A large number of writers appear not to be aware of the main objection. Instead, they argue for or against the truth of some actual or suggested premiss (e.g., (5), A , or B).¹³ See e.g., Copleston,¹⁴ Martin,¹⁵ Medlin.¹⁶

Brown says:¹⁷ "According to Aquinas, who follows Aristotle, the destructible is at some time actually destroyed."¹⁸ He then continues:

But now, the Third Way continues, if there existed *only* such inherently deteriorating things, then

¹³ Martin and Medlin reject the thesis:

C . All logical possibilities are actualized in infinite time.

Others have claimed that C is what Aquinas had in mind, and that with C supplied his argument is sound. This is not so. C can be used only because it is supposed to entail (5) or A . But the crucial question is not what entails (5) or A , but what is entailed by them. So nothing is gained by defending C . Admittedly, if (5) is rejected, the Third Way is *ipso facto* rejected what I want to show is that the Third Way fails whether (5) is rejected or not.

¹⁴ F. C. Copleston, *Aquinas* (London, Penguin Books, 1955), pp. 119–120.

¹⁵ C. B. Martin, *Religious Belief* (Ithaca, Cornell University Press, 1959), p. 151.

¹⁶ Brian Medlin, "The Contingency Argument," *Sophia* (Melbourne), vol. 5, No. 3, p. 17.

¹⁷ Patterson Brown, "St Thomas' Doctrine of Necessary Being," *op. cit.*, p. 87.

¹⁸ C. J. F. Williams has shown that Aristotle's arguments for this principle are rather confused. Although this is a very interesting issue in its own right, there is no need to go into it when discussing the Third Way.

¹⁹ E. L. Mascall, *He Who Is* (London, Longmans, Green, & Co., 1943), p. 48.

obviously the day would come in which each contingent thing had completed its own built-in process of corruption; and so on that day nothing would remain in existence. Thus, to take a simple analogy, if clocks ceased existing as soon as they ran down, then the time would come when there were no clocks left—unless we also hypothesize some being continually to wind them and/or make new clocks. . . . Moreover, if past time be infinite (which, Aquinas says, it would have to be if there were no creator, since the world cannot have come into existence without a cause), then this day of universal decay and death would already have arrived; for in an infinite amount of time all the individual processes of corruption would have been completed.

The error which I have pointed out above (in I.4) occurs where Brown says: ". . . then obviously the day would come in which each contingent thing had completed its own built-in process of corruption." Brown here confuses

some day every contingent thing has ceased to exist

with

one day every contingent thing has ceased to exist,

or, semi-formalized:

$(x)(\exists t)(x \text{ is contingent} \supset x \text{ does not exist at } t)$.

with

$(\exists t)(x)(x \text{ is contingent} \supset x \text{ does not exist at } t)$.

Another objection would be that corruption is not the same as annihilation. Even if everything that exists now had decayed beyond recognition at a time t in the future, there might still be a big heap of refuse then.

II.4 AN ATTEMPT TO REINTERPRET THE THIRD WAY

Mascall says:¹⁹

It must further be noted that the argument does not assert that, if *every individual existent* comes into being

and passes away, there must have been a moment when none of them existed. If we admit the possibility of an actually infinite manifold of things (though, as we have seen, St. Thomas would have been reluctant to do this; cf. *S. Theol.*, I, vii, 4), and suppose that No. 1 has existed for one year, No. 2 for two years, and generally No. n for n years, then each of them will have had a beginning and yet there will never have been a moment when none of them existed. What is asserted is that, if existent-being-as-a-whole came into being and passed away, there must have been a moment when nothing existed; this will still be true if we suppose existent-being-as-a-whole to have different compositions at different times. The upshot is thus that, at any moment, the existence of *something* is necessary, and the argument goes on from this.

It is also important to notice the force of the second step in the argument. Having decided that at any moment the existence of *some* thing is necessary, we want to know whether this particular thing exists just because something must exist and this happens to be the thing that does, or whether it exists because its existence is intrinsically necessary.

This is unsatisfactory. "The upshot is thus that, at any moment, the existence of something is necessary," says Mascall, and refers later to "this particular thing." But the upshot is *not* that, at any given moment, the existence of *some particular thing* is necessary, it is merely that, at any given moment, the existence of *something or other*²⁰ is necessary. Mascall wrongly infers from

$N(a \text{ exists or } b \text{ exists or } c \text{ exists or } \dots)$

to

$N(a \text{ exists}) \text{ or } N(b \text{ exists}) \text{ or } N(c \text{ exists}) \text{ or } \dots$ ²¹

But apart from this flaw, there is in Mascall an interesting suggestion worth considering: that we must consider the notion of existent-being-as-a-whole to understand the Third Way.

P. T. Geach (who accepts the Third Way), apart from neglecting the main objection (of I.4) and spending some time on defending (5), says:²²

What is in fact essential to the "Five Ways" is something tantamount to treating the world as a great big object. . . . We may read the third "way" as follows: Some things are genuinely liable to cease existing. But not every thing can be of this character: for then, Aquinas tacitly assumes, a universe entirely

composed of perishable things would itself be perishable. [At this step there comes in the "lumping together" previously discussed.] Now such a universe cannot always have existed. . . .

This is unsatisfactory. Geach is confused about what he calls the "lumping together." It is introduced by him as meaning simply that the world is regarded as one big object, which may share some characteristics with other objects. The "lumping together" therefore does not entitle one to infer that the lump has the same properties, e.g., perishability, which its components have.

But apart from this flaw, there is in Geach a suggestion worth considering: that we must consider the notion of the world to understand the Third Way.

I shall now present an argument in which the suggestion in Mascall and Geach is developed. The crucial step is the step 2*-3*. The justification for this step could be stated thus: What holds universally also holds for the universe. This principle does not entail the fallacious principles of inference discussed previously in this paper (e.g., the principle: What holds for the parts also holds for the whole).

- | | | |
|----|--|---|
| 1* | What need not be at some time is not | (5) |
| 2* | Nothing need be | Hypothesis (identical with the antecedent of (6)) |
| 3* | The world need not be | 2*, instantiation |
| 4* | At some time the world is not | 1*, 3*, <i>Modus ponens</i> |
| 5* | If at some time the world is not, then at that time nothing is | Definition of "world" (?) |
| 6* | At some time nothing is | 4*, 5*, <i>Modus ponens</i> |
| 7* | If nothing need be, then at some time nothing is | 2*-6*, conditional proof |

At this stage, some further premisses must be supplied, to cater for the step from "is" to "was,"²³ and then we get:

- 8* If nothing need be, then at some time nothing was.
8* is identical with (6).

²⁰ For a discussion of this distinction, see P. D. Shaw, "On the Logic of Particularity-Assumptions," *Mind*, vol. 77 (1968) pp. 186-205. The upshot, assuming certain versions of *ex nihilo nihil*, and assuming that something exists now, is therefore only that there was always something or other. This may be phrased: there was always the world. This does not suffice for holding: the world exists necessarily. But in any case, "the world exists necessarily" would be ill suited to back up any cosmological argument for the existence of God.

²¹ Cf. $N(p \vee \sim p)$, which is true, and $(Np \vee N\sim p)$ which is false, for all relevant interpretations of ' N '.

²² G. E. M. Anscombe and P. T. Geach, *Three Philosophers* (Oxford, Blackwell, 1963), pp. 112-115.

²³ See above Sect. I.4.

Is this what Aquinas had in mind when he moved from (5) to (6)? I doubt it. Nothing in his text warrants such an interpretation. But this question need not concern us: the argument above is unacceptable anyway.

Semi-formalized, the step 2^*-3^* would read:

$F_2^* (\forall x) \sim (x \text{ need be})$

$F_3^* \sim (W \text{ need be})$

Call the things which instantiate ' x ' a, b, c, \dots

If W , as Mascall suggests, is existent-being-as-a-whole, W must contain a, b, c, \dots . W is not necessarily identical with any of a, b, c, \dots ; therefore, we are not entitled to infer $3^*,^{24} 5^*$, on the other hand, is true. If W , as Geach suggests, is *not* existent-being-as-a-whole, but a (big) object, apart from which something else may exist, then we are entitled to infer 3^* , but now 5^* is false. So either 2^*-3^* is valid, and 5^* false, or 2^*-3^* is invalid and 5^* true. Therefore this argument fails.

Australian National University

Received September 24, 1968

²⁴ In general, if x ranges over individuals in a universe of discourse u , then the argument form " $(\forall x)(\phi x)$, therefore ϕu " is invalid. For there is one argument of that form which is invalid: let u be objects which weigh 1 kg., and let ϕ stand for the predicate "weighs 1 kg." Thus interpreted, $(\forall x)(\phi x)$ is true, but ϕu is false.

There can of course be interpretations of ϕ and u such that both $(\forall x)(\phi x)$ and ϕu are true. But for our discussion only the validity question is relevant.

V. ON "INTENTIONALITY" AND THE PSYCHOLOGICAL

W. GREGORY LYCAN

SEVERAL philosophers have recently tried, following Brentano, to establish the claims that (1) a logical property called "intentionality" is an ineluctable characteristic of all sentences containing "psychological" verbs, and that (2) no non-psychological verb is "intentional." Statements about propositional attitudes receive special consideration in these investigations.

What initially renders evaluation of their results difficult is that no less than nine differently specified concepts of "intentionality" are being appealed to, some of which are restricted to sentences of certain logical types. If we are to make sense of the above two claims, we must sort these out as rigorously as possible. But first let me set out the exact nature of the enterprise more explicitly, so that the grounds upon which some of the definitions are to be rejected will be absolutely clear: We want to find a definition of "intentionality" *such that* it will be a contingent (or in some sense non-trivial) fact that (1) and (2) above obtain. A criterion put forth as logically necessary and sufficient of intentionality may be rejected, therefore, if some psychological verb fails to count as intentional according to it, or if it is fulfilled by some obviously non-psychological verb. In such a case, the recalcitrant verb will be referred to as a *counterexample*. But it must be emphasized that rejections are to be made entirely on the basis of *usefulness* vis à vis the vindication of what I shall hereafter call Brentano's thesis;¹ I shall never mean to suggest that a proposed *definition* admits of a truth-value, although I shall insist that a definition whose adoption would lead to the falsification of Brentano's claims be discarded. (My purpose is to see *if there is* any logical property for which (1) and (2) are correct.)

I think it is most fruitful to stipulate at the outset that the adjective "intentional" shall apply in its primary usage to *verbs*, and then to say that a

sentence is intentional if and only if one of its verbs is intentional. (Later on I shall liberalize this recommendation slightly.) Now we can make the preliminary distinction between verbs which take propositions as their objects and those which take entities given by names or by descriptions. Clearly there are psychological verbs of each of these types; so a useful concept of intentionality must apply to both, although I shall occupy myself primarily with investigating the logical properties of the former sort. Let us hereafter speak of "proposition-verbs" ("*p*-verbs") and "object-verbs" ("*o*-verbs"). Some verbs can, of course, be both *p*-verbs and *o*-verbs.

I can now list the most interesting of the sets of defining conditions offered. (Some of the nine, like (B), are not "defining" conditions in the most precise sense, since they are not put forth as both necessary and sufficient, and hence are not intended to be strictly constitutive of intentionality. But they are at least fragmentarily suggestive ways of marking off the concept.) I shall restrict my discussion in this paper to those sets of criteria to which there are no *obvious* counterexamples, or at least to those sets which have yet to be controverted in the literature. In recording them as suggestions, however, I do not mean to say that their original proponents still adhere to them.

(A) In "The Intentionality of Sensation: A Grammatical Feature,"² G. E. M. Anscombe loosely describes an intentional *o*-verb as one which satisfies three conditions; they are jointly sufficient, although she does not say that the first, at least, is necessary:

- (i) The actual existence of the referent of the direct object of the verb is not a necessary condition of the truth of the clause in which the verb occurs.

¹ Brentano, of course, did not formulate (1) and (2) in terms of "logical" or "grammatical" properties (see his *Psychologie vom empirischen Standpunkt* [Vienna, 1874]). The connection between his "psychological theory of intentionality" and the view that I am evaluating in this paper is best expressed by Roderick Chisholm in his article, "Intentionality" in Paul Edwards (ed.), *The Encyclopedia of Philosophy* (New York, Macmillan, 1967).

² G. E. M. Anscombe, "The Intentionality of Sensation: A Grammatical Feature" in R. J. Butler (ed.) *Analytical Philosophy: Second Series* (Oxford, Basil Blackwell, 1965).

- (ii) The substitution of a different description of the object can affect the truth-value of the original clause.
- (iii) The description of the object may be "indeterminate," in that certain questions which might sensibly be asked of an object failing to meet (i) and (ii) need not be applicable in the intentional case. ("I can think of a man without thinking of a man of any particular height [intentional *o*-verb]; I cannot hit a man without hitting a man of some particular height [non-intentional *o*-verb], because there is no such thing as a man of no particular height.")

I talk of "clauses" rather than of sentences in my formulation to prevent us from saying, e.g., that "I kicked a rock or I punched Fred" meets (i) because I may have kicked a rock even if there is no such thing as Fred.

This definition is the only one of the nine that applies exclusively to *o*-verbs. It can, I think, easily be translated into *p*-verb language; I shall attempt to do so below.

(B) In "Sentences About Believing"³ Roderick Chisholm offers three conditions of intentionality for whole sentences (the first two appear in *Perceiving*⁴ as well). He apparently means them to be individually sufficient, but none of them taken singly is necessary. Thus he gives us no way of saying that a particular sentence is *not* intentional, unless we assume that the disjunction of the sufficient conditions is actually a necessary condition. The first of the three is similar to (A-i), Anscombe's nonexistence condition for *o*-verbs. The other two apply to sentences containing *p*-verbs: When a sentence *S* contains a *p*-verb having *q* as its object, *S* is intentional if

- (ii) neither *S* nor its contradictory implies either the truth or the falsity of *q*;
- or (iii) *q* contains a name or description which (in Frege's terminology) has an "indirect reference."

The remaining definitions presuppose a notation devised by Chisholm in "On Some Psychological Concepts and the 'Logic' of Intentionality."⁵ He introduces the idea of a "modal prefix," which is

"a phrase ending with 'that,' or at the end of which the word 'that' may be inserted"; it serves as a sentence-forming operator, applicable just to sentences. He now specifies four kinds of sentences which can be formed by combining a given modal prefix "It is *M* that" with a universal or an existential quantifier. They are

UC: "It is *M* that (*x*) (*Fx*)"

UD: "(*x*) (It is *M* that *Fx*)"

EC: "It is *M* that ($\exists x$) (*Fx*)"

ED: "($\exists x$) (It is *M* that *Fx*)"

For any modal prefix, Chisholm says, a set of four sentences of this form can be constructed, and various entailment-relations can be found to hold between its members. We can represent the pattern of entailments by a 4 x 4 matrix as shown:

	UC	UD	EC	ED
UC				
UD				
EC		b		
ED				

(A dash is placed in box *b*, for example, if UD does *not* entail EC.) Every modal prefix "It is *M* that" determines such a matrix, which holds for any *Fx* whatsoever.

Let us now speak of *entire prefixes* as being intentional, since any *p*-verb can be made into a prefix without much trouble (but some prefixes cannot easily be paraphrased into verb phrases).

(C) Now Chisholm defines an intentional modal prefix as one whose matrix is

	UC	UD	EC	ED
UC		—	—	—
UD	—		—	—
EC				
ED			—	

PATTERN I

³ Chisholm, "Sentences About Believing" reprinted in H. Feigl *et al.* (eds.), *Minnesota Studies in the Philosophy of Science*, vol. II, (Minneapolis, University of Minnesota Press, 1958).

⁴ Chisholm, *Perceiving* (Ithaca, Cornell University Press, 1957).

⁵ Chisholm, "On Some Psychological Concepts and the 'Logic' of Intentionality" in H. N. Castañeda (ed.), *Intentionality, Minds, and Perception* (Detroit, Wayne State University Press, 1966); includes comments by Robert Sleight and a rejoinder by R. M. Chisholm.

or one which conforms to Pattern I when prefixed by "it is false that."

(D) For reasons which I shall discuss below, Robert Sleigh counters (C) with the suggestion that a better set of defining conditions might be:

	UC	UD	EC	ED
UC		—	—	—
UD	—		—	—
EC	—	—		—
ED	—		—	

PATTERN II (the Minimum Pattern)

(E) Finally, Chisholm proposes that a sentence prefix might be called intentional if and only if "for every sentence q , the result of modifying q by M is logically contingent."⁶

I should like to examine (C) and (D) first, since the Chisholm-Sleigh dispute centers around these despite its shotgun-like mode of attack. Chisholm obtains Pattern I for " S believes that," " S hopes that," " S fears that," etc., by presupposing that S is a "rational being," i.e., a being who believes everything entailed by each of his beliefs, hopes everything entailed by each of his hopes, etc. Sleigh claims that this assumption, when formulated explicitly (we can change " S believes that" to " S is rational with respect to his beliefs and he believes that"), is "utterly *ad hoc* and, perhaps, circular," since such a change does not constitute an analysis or paraphrase. Certainly to alter the prefix in this way would be to render Chisholm's claims about *belief* much less interesting; it might be said that they would not really be about our garden variety of belief at all, but rather about "rational opinion" or some such, since it is true and importantly so that one *can* fail to believe a deductive consequence of one's beliefs. (In fact, it is not senseless to suppose that some one of my beliefs might happen to be *incompatible* with a distant consequence of some other set of beliefs held by me. One need grant only that at least one of the beliefs in question be false.) If we are to represent the facts about our ordinary use of psychological

verbs in filling out Chisholm's matrices, then we must agree that Sleigh is correct. His "rationality" criticism holds for belief, desire, hope, fear, and wondering, to name a few; I conclude that definition (C) can be dispensed with.

In the light of this, it seems that most psychological verbs (all, I think, save perhaps "knows," "is aware," and the like) tend more to conform to the Minimum Pattern (II), as Sleigh says, and not to Pattern I or (see note 6) to Pattern III. Let us then see if we can fare any better by adopting recommendation (D).

Sleigh himself claims to show that Pattern II is not manifested only by psychological prefixes; he offers "there is a sentence derivable in the system explained on p. 32 of *Perceiving* which expresses the proposition that" as a counterexample. Chisholm replies, in essence, that "Sentence S expresses the proposition that s " can be analyzed only as "Certain people use S , or may use S , in order to assert, express, or otherwise convey the proposition that s ." To be as fair as possible to him, we might further paraphrase this latter sentence as " S is or may be used by a person when this person is thinking that s " or as " S is or may be used by a person when this person wishes to cause someone to entertain the proposition that s ." Now it may be a fact about the world that, if persons did not occasionally have thoughts and express them by creating physical (public) sentence-tokens, we would not have the idea of what it is for sentence S to express ' s '. But it is not certain that we must for this reason restrict ourselves to the analyses I attribute to Chisholm above. Sleigh's counterexample may materially imply some intentional sentence as a matter of contingent fact, but I think we should require a stronger logical connection than this if we intend to use it as grounds for saying that "semantic statements, as such, are intentional" or psychological. If we are to make a case for semantical sentences being psychological, then it would be nice to have a connection which could rule out the possibility that computers and robots may use S to express ' s ' just as persons do. One might want to say that, if something P "uses S to express ' s '" (as opposed to merely uttering S), then P is surely a person, no matter what sort of

⁶ A sixth definition, first proposed by Russell, is discussed and discarded by Anthony Kenny in *Action, Emotion and Will* (London, Routledge & Kegan Paul, 1963), pp. 199–201. A seventh and eighth are to be found in the aforementioned Chisholm-Sleigh symposium: On p. 49 Chisholm advances a new matrix-pattern (III) which includes the condition that UD and ED both entail EC. But it is easy to see that few, if any, psychological verbs meet this requirement, whether or not conformity to Pattern III is a sufficient criterion. The eighth proposal appears on pp. 50–51; it is clearly violated by " S knows that" and " S is having the occurrent thought that," and fulfilled by "It is legally obligatory somewhere in the world that."

electronic devices it may have for viscera. But this would be only a stipulation, not an argument.

There is some question, moreover, as to whether or not conformity to the Minimum Pattern could be a necessary criterion, even if it were a sufficient one, since "knowing" appears to violate it. Chisholm may be right in saying that UD entails EC for "*S* knows that," if we accept the view that "*S* knows that *p*" entails *p*. It seems justifiable to hold that, if $(x) (S \text{ knows that } Fx)$, then *S* must know that *Fa* and hence that $(\exists x) (Fx)$. Since Chisholm assumes his universe to be nonempty, there is no problem about fictional existence here.

But it might be objected that to maintain this is to presuppose that *S* is a "rational" knower. Even though the inference from "*Fa*" to " $(\exists x) (Fx)$ " seems so automatic as to be called an "inference" only by courtesy, it is still (technically speaking) an inference; and there is no difference in principle between permitting *S* the leap from "*Fa*" to " $(\exists x) (Fx)$ " and insisting that, if *S* knows the Peano axioms, he knows every theorem of arithmetic. So I shall leave it an open question whether or not our concept of "knowing" precludes our conceiving of someone's knowing that *a*, *b*, *c*, and everything else are *F* but not knowing that there are any *F* things.

What, then, does (D) leave us with? At most it is of dubious usefulness, since it is not sufficient and only *may* be necessary. (C) and (D) comprehend, in any case, only small sets of psychological prefixes which Chisholm and Sleight have selected themselves. The Chisholm matrix thus seems at best only the beginning of what might, if radically expanded, prove to be a useful definition of "intentionality." For now it is more fruitful to explore some of the other more general proposals. Eventually we hope to find an idea of intentionality which applies, not only to all psychological *p*-verbs, but to all psychological *o*-verbs as well.

(E) seems at the outset to fail for most psychological *p*-verbs: "If *S* is not having any occurrent thought, then *S* is not having the occurrent thought that" forms a logically necessary statement when completed by *any* value of *p*. Another family of counterexamples is provided by sentences such as "If *S* knows that snow is white, then it is true that *p*," which is logically necessary for at least one value of *p*, and logically contradictory for

several others, so long as we treat the incompatibility of colors as a logical matter. Nor can (E) serve as a sufficient condition: "It is legally obligatory that *p*" is logically contingent for all values of *p* (unless one holds the unpalatable view that "It is legally obligatory that *q* & not-*q*" is not merely false but nonsensical).

It will be noticed, on the other hand, that all my objections to calling (E) a *necessary* condition of intentionality turn on counter examples which contain truth-functional particles. In one of his other essays⁷ Chisholm offers a way of undercutting these criticisms. He suggests restricting (E) by allowing it to apply only to "simple" modal prefixes, a "simple" prefix being one which "contains no proper part that is logically equivalent to a sentence or to a sentence-function." Now even with this modification, (E) is still violated by "*S* knows that," "*S* notices that," etc., since "*S* knows that *q* & not-*p*" entails a contradiction. But if Chisholm were, further, definitionally to prohibit *p*'s being logically incoherent, I think he would have a useful necessary (but not a sufficient) criterion of intentionality such that Brentano's claim (1) is vindicated. At any rate, I can think of no counterexamples to this.

It is now time to examine (A) and (B). I have earlier stated my hope that (A) can be generalized to apply to *p*-verbs as well as to *o*-verbs; I shall now put forward analogues which should accomplish this.

(A') A *p*-verb or a modal prefix "It is *M* that" is intentional if and only if the following three conditions are satisfied:

- (i) For every argument *p* which is not logically incoherent, *p* entails neither "It is *M* that *p*" nor "It is false that it is *M* that *p*."
- (ii) The conjunction of "It is *M* that *p*" and "*p* and *q* are logically equivalent" does not entail "It is *M* that *q*."

((ii) is closely related to but not the same as Chisholm's condition (B-iii).)⁸

- (iii) EC does not entail ED and UC does not entail UD.

This definition is composed of a variant of proposal (B) plus the condition (A'-iii). It is easy to show that (A'-iii) is needed, since "It is legally obligatory somewhere in the world that" fulfills (B)

⁷ Chisholm, "Believing and Intentionality: A Reply to Mr. Luce and Mr. Sleight," *Philosophy and Phenomenological Research*, vol. 25 (1964-65), pp. 266-269.

⁸ The ninth (traditional) criterion of intentionality is simply failure of substitutivity, or "referential opacity." But, as Quine says, alethic contexts are referentially opaque; and further objections to the sufficiency of (A'-ii) alone are raised in this paper.

but is non-psychological. D. J. O'Connor has discussed (B) at some length,⁹ so I shall not belabor his points. But I do not think he has touched upon this particular difficulty. I find it important chiefly because, in that it centers around (B-iii), it may threaten (A'-ii) as well. Shortly I shall investigate this in detail; but for now, I think, the valuable aspects of (B) may be left to stand or fall with (A').

And (A') seems to be met by all psychological *p*-verbs, including "knows"; despite the fact that "*S* knows that *p*" entails *p*, not-*p* does not entail that *S* either does or does not know that not-*p*. So, until someone provides a counterexample, it seems that we have three adequate necessary conditions of intentionality. Are they jointly sufficient? Let us test them with Chisholm's examples of non-psychological prefixes, starting with the alethic modalities "It is possible that" violates condition (A'-i), because *p* entails "It is possible that *p*," and (A'-iii), because "It is possible that (*x*) (*Fx*)" entails "(*x*) (It is possible that *Fx*)." "It is necessary that" violates (A'-iii) similarly; and "It is logically impossible that" violates (A'-i), since *p* entails "It is not logically impossible that *p*."

What of causal modalities? Neither "*S* brings it about that" nor "*S* does something causally sufficient to prevent that" meets (A'-iii). Both satisfy (A'-ii), moreover, only when a psychological verb appears in the argument *p*; in this case we merely have a *new* prefix "It is *M* that it is *M*' that," and it should not surprise us that this latter prefix should be intentional when *M*' contains a psychological verb.

The aforementioned difficulty for (B-iii) arises when we consider prefixes of obligation. "It is legally obligatory that" does not violate (B-iii): Examine the case in which *p* takes the value "The number of tons of food produced next year shall be greater than four billion." Suppose we produce five billion tons of food next year. Then, in Frege's sense, "the number of tons of food produced next year" and "five billion" denote the same (arithmetic) object. But "It is legally obligatory that the number of tons of food produced next year shall be greater than four billion" does not entail "It is legally obligatory that five billion be greater than four billion," which is (I presume) false; (B-iii) is met, and (B) is therefore not sufficient.

Is (A'-ii) susceptible to this objection? Clearly not, for these values of *p* and *q*. (A'-ii) requires that *p* and *q* be logically equivalent, i.e., that they strictly imply each other; and "Five billion is greater than four billion" does not strictly imply anything about tons of produce save perhaps for "Five billion tons is greater than four billion tons." Perhaps a problem will appear if we use an example involving undisputably real proper names. Does "Cicero should be provided with an armed guard at all times" entail "Tully should be provided with an armed guard at all times"? By presupposing Leibniz' Law and a certain theory of naming, we can answer in the affirmative. If we do so, then it seems reasonable to say that substitutivity holds in the "legal obligation" context—what is obligatory for Cicero is obligatory for Tully, so far as the law is concerned. But I think we are still allowed to claim, denying parity of reasoning, that substitutivity fails for "*S* knows that," "*S* hopes that," etc., since *S* may (obviously) not have heard of Tully, though he knows some facts about Cicero. The difference is that a law book is not the sort of thing that is generally said to have "heard of" anyone; by its very nature it presupposes that all the empirical facts of a case (e.g., that "Cicero" and "Tully" name the same person) are known. Even if the judge who *applies* the legal canons of obligation does not know such truths, this is no concern of the impersonal law book.

I realize that what I have just offered is a somewhat artificial, naive, and dogmatic theoretical model of how legal obligation works. But it does not seem outrageous as a tentative paradigm; it gives us reasonable grounds for saying, in my opinion, that "It is legally obligatory that" violates (A'-ii), even if we permit the aforementioned Cicero-Tully entailment to go through. And if we deny Leibniz' Law or the theory of naming on which the entailment-claim rests, then (A'-ii) has clearly not been met, since it requires mutual entailment between *p* and *q*. So I conclude that legal obligation does not provide a counterexample to (A').

I take it that (A'-ii) also rules out semantical prefixes like the one invoked by Sleigh in his rejection of the Minimum Pattern. If "There is a sentence derivable in the system explained on p. 32 of *Perceiving* which expresses the proposition that *p*" is true, and *p* and *q* entail each other, then there

⁹ D. J. O'Connor, "Tests for Intentionality," *American Philosophical Quarterly*, vol. 4 (1967), pp. 173-178. O'Connor reads Chisholm very unsympathetically, but he does raise a number of real and formidable difficulties which prompt me to seek a better set of criteria than (B).

is a similarly derivable sentence which expresses the proposition that q . In general, semantical prefixes of the form "S means that" violate (A'-ii); if 'S' means that p , and p and q are logically equivalent, then surely 'S' means that q .¹⁰ Perhaps my claims here trade on a certain view of the relation between sentences and propositions which I have neither the space nor the inclination to defend; but, again, they do not seem *prima facie* implausible.

I conclude at this point that (A') is our leading candidate for the characteristic sought by Brentano. The reason I have taken such pains to establish this, perhaps at the cost of being slightly unfair to some of the other suggested definitions, is that (A) and (A') together cover both p -verbs and o -verbs in a coherent manner, as no other proposal does. This, as I have said earlier, is an enormous advantage, since many psychological verbs do not express propositional attitudes at all: "S is thinking of . . .," "S is dreaming of . . .," "S loves . . .," "S worships . . ." All of these are ruled intentional according to (A).

It may be difficult to give (A) a precise interpretation in some cases. For example, we ought to establish exactly how (A-ii) is to be used. Certainly there is a sense in which it would be true for Marius to say "I admire Cicero but I don't admire Tully (who's Tully?)," while at the same time it would be true for me to say that Marius admired Tully. But in this sense most non-psychological verbs satisfy (A-ii) as well: Marius could with equal sincerity deny that he had punched Tully in the nose. So mere first-person-third-person disparity will not serve as an adequate interpretation of (A-ii). Perhaps if (A-i) and (A-ii) were regarded as jointly rather than as individually necessary, (A-ii) would acquire some force.

In any case, I think that it is along the lines of (A') that we should assess Brentano's position. So far as I can tell, (A) & (A') is fulfilled by all of these psychological statements:

"S thinks that p "	"S realizes that p "
"S notices that p "	"S knows that p "
"S surmises that p "	"S requests that p "
"S believes that p "	"S fears that p "
"S doubts that p "	"S remembers that p "

"S wonders whether p "	"S thinks of X "
"S intends that p "	"S dreams of X "
"S affirms that p "	"S worships X "
"S states that p "	"S loves X "
"S wishes that p "	"S desires X "
"S hopes that p "	"S fears X "
"S dreams that p "	"S perceives X "

and by no clearly non-psychological ones.

I shall not stop to argue for the intentionality of perception-verbs according to (A). Discussion of this is most clearly set forth in Anscombe's own article. It will be noticed, however, that the above list fails to include any of an important category of mental concepts: those involving bodily sensations, such as pain, itches, tickles, etc. Sentences like "My arm hurts" or "My ear itches" do not contain objects at all, at least in the superficial grammatical sense. We have our choice of (a) holding that bodily sensations are too firmly rooted in the world of physiology to be (strictly) called "psychological" (a dubious assertion, since the mental states associated with them are at least not *logically* identical with bodily states), or (b) hoping that at some future time we shall be able to incorporate some intentionality-criterion, independent of the idea of a direct object, into our program which will cover these statements.

Supposing that we have indeed proved (1) and (2), what are we entitled to infer from them about the "mental" and the "physical"? "Absolutely nothing," says O'Connor:

The most that we could say would be that the structure of language does reflect our unsophisticated beliefs on this point. But it would have no tendency whatever to show that the distinction between mental and physical was in any way basic or that the mental could not be reduced to or explained in terms of the physical. So even if the distinction could be shown to be well founded, it would be no more than an interesting fact about language. . . . Perhaps our language merely embodies a discredited Cartesianism.¹¹

But surely the issue is not quite so simple as this. No doubt the plain man's language does embody many false dualistic presuppositions ("An idea popped into his mind"; "His body was in the classroom, but his thoughts were miles away"; "Everyone was charmed by his outer shell, but no

¹⁰ Some philosopher might be squeamish about admitting this, if he holds that (p) "Snow is white" and (q) "Snow is white and $2+2=4$ " are logically equivalent in that they strictly imply each other, but that they are not equivalent in *meaning*. If this view is correct, "S means that" fulfills (A'-ii). Let us therefore read "logically equivalent" (in (A'-ii)) as "synonymous" or as "equivalent in meaning"; on this interpretation, "S means that" no longer meets (A'-ii), and the possible objection is forestalled.

¹¹ O'Connor, *op. cit.*, pp. 174-175.

one could tell what he was really like inside," etc.). But these are examples of how the plain man is commonly misled by a picture or a metaphor, viz., by the "inner"—"outer" model rightly rejected by Wittgenstein. The logical property given by (A) & (A') does not have this same character; it does not betray a "diseased" adherence to any kind of metaphor. Rather, the necessity of its presence suggests that psychological statements perform a *function*, play a linguistic role, enable us to talk about something—which task cannot be

fulfilled by physicalistic statements. Brentano expressed this (but in the material mode) by saying that mental phenomena were distinguished by their essential "reference to a content" (and that physical items cannot "refer" to anything extrinsic to themselves in this same way). All this is terribly unclear, and it may not in the end serve to point up any irreconcilable difference between the "mental" and the "physical"; but it should be clarified and then assessed, not dismissed out of hand.¹²

University of Chicago

Received October 30, 1968

¹² I am indebted to Professors G. E. M. Anscombe, V. C. Chappell, Roderick Chisholm, and Robert Coburn for having pointed out several errors in an earlier version of this paper. They share no responsibility for any remaining mistakes.

VI. HUSSERL ON SIGNIFICATION AND OBJECT

JOHN E. ATWELL

IN the first investigation of his *Logische Untersuchungen*¹ Edmund Husserl puts forth an analysis of expression from the standpoint of mental acts. This rather unusual procedure results in a curious and interesting account of the relation between the signification (*Bedeutung*) of an expression and its object or referent (*Gegenstand*). My purpose is to examine the aspects of this account which help determine where Husserl stands with respect to the so-called referential theory of meaning (signification).² A proponent of this theory, I shall suppose, subscribes to the general statement that "for an expression to have a meaning is for it to refer to something other than itself," and to the specific view that "the meaning of an expression is that to which the expression refers."³ Now Husserl denies—at least verbally—both the specific view and the statement that an expression's referring to an object gives it a signification; hence he appears to deny *in toto* the referential theory of meaning. I shall argue, however, that he has no justification for either denial.

I

Since Husserl's *LU* has received relatively little attention, at least in most quarters, it will not be unfitting, I trust, to recall briefly a few of its more prominent themes. But I mention only those which have a direct bearing on the topic of this paper.

One of Husserl's primary concerns in the

first investigation is to distinguish the signification of an expression (or statement)⁴ from those features of expressions which are frequently, but in his view erroneously, regarded as significations. Advocates of psychologism tend to identify significations (concepts and propositions) with certain mental experiences ("acts of thinking"), as well as the relations between significations with the relations between mental experiences; consequently, they confuse necessary laws of logic with contingent laws of psychology. But no less mistaken, Husserl claims, are those philosophers who identify the signification of an expression (or statement) with that to which the expression refers, its object or referent. In full accord with his critique of psychologism, carried out in the *Prolegomena* (*LU*, I), and with his conception of "pure logic," Husserl is most anxious to insure the uniqueness and irreducibility of significations. For they alone are the "true objects" of logical inquiry (p. 2).

Nevertheless, belonging to an expression are all three of the aforementioned features: the signification, the object, and the mental experience (or "act"). Every expression has, in other words, three functions: it signifies something, it refers to something, and it manifests or announces something (*viz.*, one or more mental acts). Now Husserl recognizes that expressions which have no actual object, e.g., "the golden mountain" and "the round square," are perfectly significant, which means for him that they have signification. But to

¹ Edmund Husserl, *Logische Untersuchungen* (5th ed.; Tübingen, 1968). References in the text and subsequent footnotes are to volume II/1. The original work was published in 1900–1901, with substantial changes made in the second edition (1913) and retained in later editions.

² In reporting and criticizing Husserl's views I use "mean" and "meaning" to translate, respectively, his terms "meinen" and "Meinung." "Sense," which Husserl uses synonymously with "signification" (p. 52), I use to translate "Sinn." For readers of Frege's essay, "Über Sinn und Bedeutung" (*Zeitschrift für Philosophie und philosophische Kritik*, Band 100), it should be noted that, roughly, Frege's term "Sinn" is equivalent with Husserl's terms "Sinn" and "Bedeutung," and his term "Bedeutung" is fairly close to Husserl's term "Gegenstand." Unlike Frege, however, Husserl does not regard the "Bedeutung" of a true statement as "the true," but rather as a "truth in itself." With the publication of *Ideen I* (1913), Husserl begins to treat "Sinn" and "Bedeutung" as non-synonymous (but not in the manner of Frege): the former term is there applied to sensory objects, and the latter is reserved for logical or word signification (see sections 124–127). The notion of the "Sinn" (or "sens") of objects is taken over and stressed by M. Merleau-Ponty, *Phenomenology of Perception*, tr. by Colin Smith (London, 1962).

³ William Alston, *Philosophy of Language* (Englewood Cliffs, N. J., 1964), pp. 12–13.

⁴ Husserl does not fail to distinguish expressions and statements; for though every statement is an expression, not every expression is a statement. For our purposes, the distinction plays no crucial role.

say that an expression has no actual object is not to say that it has no reference to an "object" or that it is non-referential: "Every expression not only enunciates something, it also talks about something; it not only has its signification, it also refers to certain objects However, the object and the signification never coincide" (p. 46). Furthermore: "To use an expression with sense and to refer expressively to an object (to present the object) are one and the same. It does not matter at all whether the object exists, whether it is fictional, or even impossible" (p. 54). In virtue of the mental act which is (if degrees are proper) most essential to an expression and which accounts for its having signification, every expression has reference to an "intentional object." Every mental act has an intentional object, according to the tradition of Brentano, hence every act of signifying or every act of meaning something "expressively" has an intentional object. But by definition an intentional object need not actually exist.

For Husserl this is by no means a mysterious way of saying that the intentional object need not exist except "in the mind" or "immanently." He stresses, again and again, that it need not exist at all, or anywhere. For example:

I present the god Jupiter, that is, I have a certain presentational experience; in my consciousness there occurs a presenting-of-the-god-Jupiter. One may dissect by descriptive analysis this intentional experience in any way one pleases, but one will naturally not find anything like the god Jupiter; the "immanent," "mental" object does not belong therefore to the descriptive (real) constituency of the experience, thus it is really not immanent or mental. Of course it is also not *extra mentem*, it does not exist at all. But that does not affect the fact that the presenting-of-the-god-Jupiter is actual. . . . (P. 373.)

And so it is with meaning Jupiter "expressively," as in the statement, "Jupiter was a Greek god." To refer to something is, on Husserl's view, somewhat analogous to challenging someone to a contest. If the challenge is not taken up, one has nevertheless made a challenge; and if there is no actual object of reference, one has nevertheless made a reference or referred to an "object."

As we have just seen, it is the mental act which accounts for an expression's having a signification; thus Husserl might well quarrel with the statement (from William Alston) that "for an expression to

have a meaning is for it to refer to something other than itself"—assuming this to mean that the expression's referential character provides the expression with a signification. For Husserl claims: "An expression gets a reference to an object only because it has a signification. . . ." (P. 49.)⁵ Such a dispute would be merely verbal, however, for according to Husserl the mental act which allegedly gives the signification necessarily refers to an intentional object. He also says that "the act of signifying is the definite way of intending the object" (*Ibid.*), and he explicitly cautions against the error of distinguishing the act which signifies from the act which refers, means, or intends (p. 50). In other words, the mental act of signifying is, from a slightly different perspective, the same as the act which refers to an object.

Since Husserl analyzes expressions—perhaps better, uses of expressions—in terms of mental acts, certain parallels arise between (uses of) expressions and mental acts in general. We have remarked that an expression has reference to an object because the mental act essentially associated with the expression has reference to an object; just as every act has, in addition, a "content," so every expression has a signification. Indeed, the "content" of the act of meaning something expressively is identical with the signification of the expression used (p. 46). And just as an act's object should never be identified with its "content" (i.e. just as the object *that* is intended must be distinguished from the object *as* it is intended [p. 400]), so an expression's object should never be identified with the expression's signification. To say the former is really to say the latter. Finally, since the act's "content" is never the act itself, so the expression's signification is never the act of meaning something expressively itself. In summary: to refer to something with an expression, or expressively, is to mean (*meinen*) something; that which is meant (*das Gemeinte*, the object) is never identical with the way in which it is meant; the way of meaning something, the "mode of presentation" (in Frege's terms), is the signification of the expression; and the act of meaning something expressively, sometimes called quite understandably the "act of signifying," is distinct from both the object meant and the signification of the expression used.

What I have been sketching can be, and for the sake of clarity probably should be, related to

⁵ It is the possession of a signification which distinguishes expressions from mere signs. Husserl points out at the very beginning of the first investigation that signs, e.g., fossil bones, "signify" or "point to" something other than themselves, but they have no signification (pp. 23–25).

Husserl's radical division of existence into two realms—the ideal and the real. The real is characterized “essentially” by temporality (p. 123); everything real occurs, so to speak, within time. Real existence comprises, consequently, all mental acts, physical objects, persons, and events; in short, whatever is datable. The ideal, which includes all significations as well as “species” (or universals), consists of non-temporal or timeless entities. Most important for Husserl's interest is the distinction between real mental acts and ideal significations.⁶ When I utter, for example, “The three altitudes of a triangle intersect at one point,” my act of judging is “a fleeting experience, originating and passing away,” i.e., it is a real existent. But what I say, the proposition or *Wahrheit an sich*, “does not originate and pass away,” i.e., it is an ideal existent; it is moreover the signification of my statement (p. 44) or, to say the same thing, the “content” of my act of judging. Ideal existents are called “ideal unities,” for they are unities in multiplicities. For instance, the signification of an expression stands to the multiple actual and possible acts of expressing that particular signification (concept or proposition) as the universal redness stands to its multiple instances, i.e., as the One stands to the Many (pp. 100, 106).⁷ Husserl claims that the signification of an expression or statement is ideal, no matter whether the expression's object is real or ideal, and no matter what kind of expression is uttered.

II

One might take issue, of course, with a number of Husserl's views on ontology and the philosophy of language, but I intend to consider only two contentions both of which relate to the referential theory of meaning: one, that every expression refers to an object; two, that the signification of an expression and its object never coincide. Most current philosophers will concur with Husserl's second contention, but they will reject the first one.

It is currently quite common to distinguish between language and speech, between type and

tokens, and—most relevant to the issue at hand—between linguistic units and uses or utterances of linguistic units. Earlier I mentioned that Husserl is attempting to analyze uses of expressions or, as he often puts it (p. 32), expressions in their communicative function. But, failing to observe the above distinctions, he sometimes talks as if every linguistic unit or “word” had reference to an object. That this is not (or at least should not be) his view becomes clear upon realizing what he takes an expression to be: namely, a “physical object” such as a mark on paper, a sound, etc., which someone produces with the intention of putting forth some “thought” or of communicating something to someone (p. 31). “In itself” an expression is a physical object like any other (p. 407), but in virtue of the mental act of meaning something with it or of understanding something by it, the physical “expression” becomes a genuine expression.

An articulated complex of sounds (or written marks, or the like) first becomes a spoken word or communicative speech through the fact that the speaker produces it with the intention of “expressing [*äussern*] himself about something” thereby; or in other words, in certain mental acts the speaker gives to it [the complex] a sense which he wants to communicate to the hearer. (P. 32.)⁸

Spots of ink and sounds “mean” nothing until persons, in virtue of some mental activity, “mean” something with them or “understand” something by them. Husserl analyzes the act of meaning something with an “expression” in the same way that he analyzes the act of understanding an “expression” (p. 74, n4); both are characterized by “going beyond” what is produced or what is perceived to, as he would say, what is meant.

Mental activity—if it may be so called—consists of intentional acts, and on Husserl's conception of such acts some object is necessarily intended or referred to. That every expression has an object means then that every case of expressing (i.e., *jedes Ausdrücken* rather than *jeder Ausdruck*) has reference to some object, or that every act of meaning “means” some object. Quite clearly, “object” must be understood as covering anything

⁶ For Gilbert Ryle's account of what I say just below, see his “The Theory of Meaning” in *Philosophy and Ordinary Language*, ed. by Charles E. Caton (Urbana, 1963), p. 149.

⁷ For a discussion of Husserl's ontology, see Gustav Bergmann, “The Ontology of Edmund Husserl” in the author's *Logic and Reality* (Madison, 1964), pp. 193–224.

⁸ Bergmann seems to agree: “Words spoken are sounds among sounds; words written, marks of chalk or ink. By themselves, they do not represent anything, don't refer to anything. We refer by means of them to what we have made them represent,” “Meaning” in *ibid.*, p. 93.

mentionable or referable. Whenever a person expresses himself, he expresses himself *about* something and he gives expression *to* something, e.g., a judgment, wish, command, or the like. That which he expresses himself about is the object of his expression, and that which he gives expression to (or manifests, *kundgeben*) is some mental act.⁹ Advocates of the "intentionality thesis of mind" tell us that a priori if I judge, I judge (about) something; if I wish, I wish for something; if I ask, I ask about something; and so on. In short, Husserl regards the claim that every expression has an object as a priori. Consequently, there is absolutely no problem in understanding his affinity to one aspect of the referential theory of meaning: it follows rigorously from his notion of an expression and his conception of mental acts.¹⁰

Preferring a sympathetic interpretation to an antagonistic one, I have tried to show that Husserl's claim that every expression has an object is at least not wildly absurd. He would surely deny, for instance, that in saying "This book is red," I am executing four mental acts and referring thereby to four distinct objects. He would maintain, rather, that every "word" is used either to refer to something or else in conjunction with other words to make a reference. Even on this most charitable interpretation, his view may be unsound, but it is not obviously unsound: one cannot refute it by simply pointing out that "or" refers to no object. It must be admitted, however, that Husserl does not follow through with the strand in his thought upon which my interpretation rests. Had he done so, and had he pursued further the multifunctionality of expressions, he would surely have arrived at a position not unlike that of several current philosophers. One need not stretch a point too extremely to see a number of similarities between Husserl's various "acts of expressing" and J. L. Austin's various "speech acts."¹¹

III

J. N. Mohanty, in his recent book on *Edmund Husserl's Theory of Meaning*,¹² goes to considerable length to show that Husserl rejects the referential

theory, i.e., that Husserl refuses to identify the signification and the object of an expression. No doubt Husserl does "reject" the referential theory (this we have seen), but whether he is justified in so doing is another question. It is not unimportant to take note of the following: (1) Husserl admits that cases arise in which a change in the meant object effects a change in the signification of "the same words" (pp. 79-80); (2) he insists that the act of signifying is not to be distinguished from the act of meaning or intending an object (p. 50); (3) he grants that it is impossible to describe acts of meaning (*meinende Akte*) without having recourse to the meant things (*die gemeinten Sachen*) (p. 11)—which seems to say that we cannot distinguish, e.g., the act of meaning *that chair* from the act of meaning *this book* without talking about the chair and the book; and (4) following his long discussion of expressions which fluctuate in signification from case to case, viz., "occasional" or "indexical" expressions, and of expressions which "do not possess an identical signification-content in every case of their application" (p. 88), e.g., empirical expressions like "horse," Husserl finally allows that the only expressions free from fluctuation of signification are exact expressions, i.e., those which occur in propositions of mathematics and pure logic and which allegedly refer to ideal, immutable objects and states of affairs. In light of these remarks, and others, I propose to argue that Husserl has no justification for rejecting the referential theory of meaning.

The main arguments for distinguishing signification and object occur in Sect. 12 of the first investigation where Husserl calls attention to (1) expressions which have different significations but the same object, and (2) expressions which have different objects but the same signification (pp. 47-48).¹³ In illustration of (1) he mentions the following "names": "the Victor of Jena" and "the Vanquished of Waterloo." Husserl does not say that these two expressions refer to Napoleon, though presumably he thinks that they do; and he does not argue that they have different significations, stating simply that they obviously (*offenbar*) do. If expressions "*E*₁" and "*E*₂" refer

⁹ In cases other than mere judgment or assertion, one gives expression to or manifests more than one act; for example, in saying "I wish you luck," one manifests an act of wishing and an act of signifying.

¹⁰ I mention this because several writers express some puzzlement over their predecessors' subscription to the referential theory of meaning, e.g., Ryle in "The Theory of Meaning," Caton, *loc. cit.*, pp. 131-132.

¹¹ J. L. Austin, *How to do Things with Words* (London, 1962), lecture viii and following.

¹² J. N. Mohanty, *Edmund Husserl's Theory of Meaning* (The Hague, 1964), especially pp. 17-23, 43, 75-76, 98.

¹³ Mohanty appears to accept both arguments, for he paraphrases them (p. 17) without any critical discussion.

to the same object and yet have different significations, then—the argument appears to go—the signification of “ E_1 ” (or “ E_2 ”) must be distinguished from the object of “ E_1 ” (or “ E_2 ”). The problem facing Husserl is pointed up by asking how he proposes to distinguish between significations. My contention is that he is unable to do so except by appeal to intentional objects, i.e., to the respective objects of the expressions in question. If S_1 can be distinguished from S_2 only by distinguishing O_1 from O_2 , then there appears to be no justification for distinguishing S_1 from O_1 or S_2 from O_2 .

To use an expression with sense or to understand it is, on Husserl's view, to perform an act of meaning or understanding. And such an act, like every other, has reference to an object, though the object—being an intentional object—need not actually exist. Now when we understand “the Victor of Jena” and then “the Vanquished of Waterloo,” we are—Husserl seems to hold—referentially directed to one and the same object, viz., Napoleon; but we refer to it (him) in different “ways,” first as the Victor of Jena and then as the Vanquished of Waterloo. Alternatively stated, the two acts of understanding have the same object but different “contents.”

Notice, however, that it is solely in virtue of a fair bit of *knowledge* about French history that we are able to provide this account. Suppose someone knew practically nothing about French history. He would no doubt understand the two expressions, and he would recognize—in spite of his ignorance—that they differ in signification. By the mere understanding of these expressions, this person would be “conscious” of different significations. “Whenever we use an expression or understand it, it signifies something for us, i.e., we are actually conscious of its sense” (p. 183). But what is he “actually conscious” of in understanding the two expressions? What he is referentially directed to? Obviously, such a person is unable to give different answers to these questions, i.e., he is unable to distinguish what he is conscious of (allegedly, the senses) from what he is referentially directed to (allegedly, the object). In short, knowledge and never mere understanding of an expression allows one to distinguish the signification from the object. But Husserl holds that in merely understanding an expression one is referentially directed to an object; in fact, to be referentially directed to an object is nearly the definition of understanding an expression. Hence, when I understand, e.g., “the

seventh man to climb Mt. Everest,” I am referentially directed to an object; but the only object I am able to cite is *the seventh man to climb Mt. Everest*.

What can it mean to say, on Husserl's view, that the expression “the Victor of Jena” refers to Napoleon? Clearly not that these marks on paper “mean” (refer to) Napoleon, for until animated by a sense-giving act marks on paper “mean” nothing, until then they do not even constitute an expression. Nor can it mean that in order to understand the expression one must be referentially directed to (or “have in mind”) Napoleon—or, if it did mean this, it would be false—though one must be referentially directed to some object. To even make sense of the claim, we must, so to speak, go beyond understanding the expression to a consideration of what most people know, viz., that Napoleon was the Victor of Jena. But, as I have insisted, knowing this fact is not a requisite for understanding the expression, thus for being referentially directed to some object. In fact, it is just the converse. Knowing what an expression refers to, in the sense that “the Victor of Jena” refers to Napoleon, is not even relevant to an analysis of understanding the expression. But only by bringing to the analysis the extraneous feature of knowledge is Husserl able to present the least semblance of an argument for distinguishing the signification from the object.

That one signification differs from another is said to be obvious or given (p. 183); likewise, that one “act-character” differs from another is said to be “beyond description” (p. 384). These are, I think, revealing assertions. For, first, they reveal Husserl's inability to account for sameness and difference of significations; and second, they lead us to ask “What is given in an act of understanding, or any act for that matter?” The answer is, of course, the intentional object; nothing is more evident than the object of a mental act. In understanding “the Victor of Jena,” the intentional object *the Victor of Jena* is given; and in understanding “the Vanquished of Waterloo,” *the Vanquished of Waterloo*. One should not forget that intentional objects can be anything. Our person ignorant of French history recognizes that the two expressions differ in signification because he recognizes that the two respective intentional objects differ. To recognize the one is to recognize the other.

Further reasons could be adduced in support of my argument, but I see no need for this. I want to turn briefly to Husserl's second “argument” for

distinguishing signification and object. It is, unfortunately, hopelessly confused, for—among other things—Husserl fails to heed the difference between an expression type and expression tokens, and he begins talking about “objective reference” instead of “object.” He says that “two expressions can have the same signification but a different objective reference” (p. 47); but he means that two uses of one expression can refer to different objects although the signification of the expression remains the same. Unlike most current philosophers, who use so-called “indexical expressions” to illustrate this claim, Husserl uses the expression “a horse.” In the statement “Bucephalus is a horse” and then in “This old nag is a horse,” the expression “a horse,” says Husserl, has the same signification; but in the former statement it refers to Bucephalus while in the latter it refers to this old nag. To begin with, one wants to say that “a horse” is not a referring expression at all, or at least that it does not refer to this, then that individual horse. If anything, it refers to the universal horseness, or what Husserl calls a “species” or “essence.” Secondly, if, in the first statement, “a horse” did refer to Bucephalus and “Bucephalus” also referred to Bucephalus, then the statement would be analogous to, say, “Napoleon is the Victor of Jena.” But of course it isn’t. And, finally, when Husserl does deal with expressions which retain the same signification in different uses, e.g., “this” and other indexicals, he claims that they change signification with each different use, i.e., with each different object or state of affairs they are used to refer to (pp. 79–80). Hence, the second “argument” proves wholly unsuccessful.

I have attempted to show that by analyzing

meaning and understanding in terms of intentional acts Husserl has no grounds for maintaining that the signification and object of an expression never coincide. Given his conception of intentional acts, Husserl may make only an analogous distinction at best—perhaps that between the direct object of an expression (which might be called the signification) and the indirect object (which might be called the object). On this view, the direct object of “the Victor of Jena” would be *the Victor of Jena*, and the indirect object would be—if there is one—the person victorious at Jena. In this example, as opposed to, say, “the golden mountain,” there is such an indirect object, viz., Napoleon, whether one knows it or not. But then, expressions which have no indirect object, e.g., “the golden mountain,” would have—using Husserl’s terminology again—no object at all; and this has been denied. Or else, the signification of such expressions, i.e., the direct object of reference, would be identical with the only object of reference; and this too has been denied. Consequently, the suggestion that by signification and object Husserl actually means direct object and indirect object helps to preserve his views solely with regard to definite descriptions which “fit” an actually existing thing or person. But Husserl wants to extend his distinction between signification and object beyond such a limited sphere. Hence the suggestion fails to extricate him from the difficulties I have pointed up. I don’t want to leave the impression, however, that Husserl has nothing of interest or importance to say about expressions, “acts,” and related issues; on the contrary, I think that for anyone concerned with such issues Husserl’s *LU* is *ein unentbehrliches Werk*.¹⁴

Temple University

Received October 30, 1968

¹⁴ The writing of this paper was supported by a Temple University Summer Research Award (1968), for which I express my gratitude.

VII. A NEGLECTED ASPECT OF THE RELATIONSHIP BETWEEN BERKELEY'S THEORY OF VISION AND HIS IMMATERIALISM

ROLF SARTORIUS

THE "design" of Berkeley's *An Essay Towards a New Theory of Vision*, according to its author,

is to show the manner wherein we perceive by sight the distance, magnitude, and situation of objects. Also to consider the difference there is betwixt the ideas of sight and touch, and whether there be any idea common to both senses.¹

An examination of Berkeley's private notebooks (now usually referred to as *The Philosophical Commentaries*), which were completed in 1708, reveals that the immaterialist doctrine for which Berkeley is best known was well developed prior to the publication of the *NTV* in 1709. And yet this doctrine is only explicitly introduced with the publication of Berkeley's *Principles of Human Knowledge* in 1710. Thus a quite natural question, one to which most commentators on Berkeley have expressly addressed themselves,² is this: What bearing, if any, does the "design" of the *NTV* have on the immaterialism of the *Principles* and the *Three Dialogues Between Hylas and Philonous* (1713)? What I shall argue here is that a significant part of the answer to this question has been overlooked by Berkeley scholars, this being due at least in part to Berkeley's own remarks in the *Principles* on the relation between his theory of vision and his immaterialism.

2. In Sects. 43 and 44 of the *Principles* Berkeley considers as a likely objection to his immaterialism the claim that we immediately perceive by sight external objects existing at a distance from us. "But for the fuller clearing of this point," he replies,

it may be worthwhile to consider how it is that we perceive distance and things placed at a distance by

sight. For that we should in truth see external space, and bodies actually existing in it, some nearer, others farther off, seems to carry with it some opposition to what has been said of their existing nowhere without the mind. The consideration of this difficulty it was that gave birth to my *Essay Towards a New Theory of Vision*, . . . wherein it is shown that distance or outness is neither immediately of itself perceived by sight, nor yet apprehended or judged of by lines and angles, or anything that has a necessary connection with it; but that it is only suggested to our thoughts by certain visible ideas and sensations attending vision, which in their own nature have no manner of similitude or relation either with distance or things placed at a distance; but by a connection taught us by experience they come to signify and suggest them to us after the same manner that words of any language suggest the ideas they are made to stand for . . .

The ideas of sight and touch make two species entirely distinct and heterogeneous. The former are marks and prognostics of the latter. That the proper objects of sight neither exist without mind, nor are the images of external things, was shown even in that treatise. Though throughout the same the contrary be supposed true of tangible objects—not that to suppose that vulgar error was necessary for establishing the notion therein laid down, but because it was beside my purpose to examine and refute it in a discourse concerning *vision*.

In brief, what Berkeley suggests here is that the *NTV* is of metaphysical significance in that (a) it contains the reply to a natural objection to his immaterialism, and (b) this reply represents an intermediate position with regard to the doctrine of the *Principles* and the *Three Dialogues* in that it commits him to an immaterialism of sight, but not of touch.

With the exception of Armstrong, who has argued that there is simply no connection at all

¹ Sect. 1. This work will henceforth be referred to as the *NTV*.

² See C. M. Turbayne, "The Influence of Berkeley's Science on His Metaphysics," *Philosophy and Phenomenological Research*, vol. 16 (1955-56), pp. 476-487, and the references therein. Turbayne writes, on p. 476, that "The history of philosophy provides few cases in which such a curious and puzzling relation holds between a man's scientific discoveries and his metaphysical theories as is found in the case of Berkeley."

between the *NTV* and Berkeley's immaterialism,³ commentators have by and large rested content with explaining and elaborating upon these passages from the *Principles*. Luce endorses the view they express and claims that it is confirmed by the *Philosophical Commentaries*;⁴ Warnock goes beyond Berkeley's own statement only in adding that one reason Berkeley may have written the *NTV* is that there was much interest in works on vision at the time at which he wrote.⁵

Even Professor Turbayne, who has explored the problem at hand at length in a number of places, does not move far beyond Berkeley's own statement:

The two important conclusions regarding vision, from the standpoint of the metaphysics of the *Principles*, . . . were . . . (1) The proper objects of sight do not exist without the mind; (2) The proper objects of sight are not the images of external things.⁶

Elsewhere, Turbayne has restated the second conclusion in the following way: "The proper objects of sight are only images of themselves." Turbayne's contention is that this is both a visual analogue to, and a confirming instance of, what he takes to be Berkeley's major metaphysical principle that "Neither our ideas nor anything like our ideas can possibly be in an unperceiving thing."⁷

3. A preoccupation with the immaterialism of sight of the *NTV* has blinded commentators to what seems to me to be an equally significant aspect of the relationship between Berkeley's theory of vision and his metaphysics. The clue is provided by Berkeley himself, who in *The Theory of Vision or Visual Language Vindicated and Explained* (1733) writes that:

It has indeed been a prevailing opinion and undoubted principle among mathematicians and philosophers that there were certain ideas common to both senses, whence arose the distinction of primary and secondary qualities. But I think it has been demonstrated that there is no such thing as a common object, as an idea, or kind of idea perceived both by sight and touch.⁸

This brief and almost casual remark, which to the best of my knowledge is unparalleled in Berkeley's writings and which as far as I know has been ignored by Berkeley scholars, suggests that Berkeley may have looked upon the *NTV* as containing a refutation of Locke's distinction between primary and secondary qualities, and thus as constituting an important step toward "ye immaterial hypothesis" of the *Principles* and the *Three Dialogues*. The thesis that there are no "common sensibles"⁹—but that the ideas of the various senses (in particular sight and touch) are both numerically and specifically distinct—was what Berkeley called "the main part and pillar"¹⁰ of his theory of vision. What I shall contend, following the passage from the *TVV*, is that Berkeley thought that Locke's distinction between primary and secondary qualities was so closely linked to the distinction between common sensibles and proper objects that a demonstration that the latter distinction was untenable would show the same for the former. For although Berkeley, of course, rejected both of these distinctions, I shall argue that he believed that within the Lockean framework all ideas of primary qualities must be accorded the status of common sensibles. The conclusion of the *NTV* that there are no common sensibles, in conjunction with this principle, entails the further conclusion that there are no ideas of primary qualities, such qualities being understood as being independent of the mind and "really in" external objects. Although the claim that there are no ideas of primary qualities is of course neither a statement of Berkeley's immaterialism nor a direct argument for it, it is an important step toward Berkeley's immaterialism in that the negation of this claim would entail the falsity of the immaterialist doctrine.

4. How, it might be asked, could Berkeley possibly have believed that all ideas of primary qualities are common sensibles when (a) the primary qualities of number and solidity are excluded by Locke from his list of common sensibles, and (b) none of Locke's various ex-

³ D. M. Armstrong, *Berkeley's Theory of Vision* (Melbourne, Melbourne University Press, 1960). See especially, ch. 2, Sect. 6.

⁴ A. A. Luce, *Berkeley and Malebranche* (Oxford, Oxford University Press, 1934), ch. 2.

⁵ G. J. Warnock, *Berkeley* (London, Pelican Books, 1953), ch. 2 and ch. 3.

⁶ C. M. Turbayne, "Berkeley and Molyneux on Retinal Images," *The Journal of the History of Ideas*, vol. 16 (1955), p. 344.

⁷ "The Influence of Berkeley's Science on His Metaphysics," *op. cit.*, p. 486.

⁸ Sect. 15. Italics mine. This work will henceforth be referred to as the *TVV*.

⁹ Aristotle, *De Anima*, Bk. II, ch. 6 (418a), writes: "Common sensibles" are movement, rest, number, figure, magnitude: these are not peculiar to any one sense, but are common to all. There are at any rate certain kinds of movement which are perceptible both by touch and sight.

¹⁰ *TVV*, Sect. 41.

planations of the distinction between primary and secondary qualities implies that all ideas of primary qualities are common sensibles? Perhaps it is because these questions have appeared to be unanswerable that writers such as Armstrong, Luce, Warnock, and Turbayne have made no connection at all between Berkeley's attack on the doctrine of common sensibles and his metaphysical concern with Locke's distinction between primary and secondary qualities.

In Bk. II, ch. V, of the *Essay*, Locke writes:

The ideas we get by more than one sense are, of *space* or *extension*, *figure*, *rest*, and *motion*. For these make perceivable impressions, both on the eyes and touch; and we can receive and convey into our minds the ideas of the extension, figure, motion, and rest of bodies, both by seeing and feeling.

And in ch. VIII, Sect. 9, of the same book:

These I call the *original* or *primary qualities* of body, which I think we may observe to produce simple ideas in us, viz., solidity, extension, figure, motion or rest, and number.

So number and solidity do appear on Locke's list of primary qualities, but not on his list of common sensibles. And there can be no question but that Berkeley, who knew the text of the *Essay* quite well, was aware of this.

5. There really is no problem with number. For in ch. XVI, Sect. 1, of Bk. II Locke indicates that he views the simple idea of number as being the common sensible *par excellence*, in that it accompanies (and thus renders complex) every other idea which we experience prior to abstraction whatsoever:

Amongst all the ideas we have, as there is none suggested to the mind by more ways, so there is none more simple, than that of *unity*, or one: it has no shadow of variety or composition in it: every object our senses are employed about; every idea in our understandings; every thought of our minds, brings this idea along with it.

The reason number is not included in Locke's list of common sensibles, I suspect, is simply because that list is one of ideas of *sense alone* (and not of reflection) which "we get by more than one sense." Berkeley, of course, was quite properly concerned with the notion of a common sensible in the broader and unrestricted sense in which number is clearly a common sensible for Locke.

6. Solidity is more difficult to deal with, for in Bk. II, ch. IV, Sect. 1, Locke writes:

The idea of solidity we receive by our touch, and it arises from the resistance which we find in body to the entrance of any other body into the place it possesses, till it has left it.

And the first section of the preceding chapter makes it clear that Locke is considering the idea of solidity as the most important simple idea which comes from the sense of touch *alone*.

The only ground which Berkeley could have had for accepting and (as I shall show he did) employing the principle that all ideas of primary qualities are common sensibles, then, is a conviction that Locke—on his own terms—was simply mistaken in not treating solidity as a common sensible. And I believe it is quite plausible to attribute this view to Berkeley, one reason being that, at least at a certain level of analysis from within a Lockean framework, it is a reasonable view.

For can't we just as well see as feel one body supporting another, or one body in motion making contact with, and failing to penetrate the surface of, another body? In short, can't we just see—as we often say we do—that a body is solid? Given the fact that Berkeley denied that there were any common sensibles at all, and that he thus in a sense had to think himself into the Lockean framework, isn't it quite possible that he may have thought that if Locke was to claim that we see and feel the same extension and figure, for instance, he should also have claimed that we see and feel the same solidity?

One might reply here that solidity is different from the other Lockean primary qualities in the crucial respect that although we could be said to *perceive immediately* motion (or rest), figure, extension, and number by either sight or touch, anything that could be properly described as the visual perception of solidity must involve *inference*. Strictly speaking, it might be said, we cannot *see* that one object is *supporting* another, but only that the one is (say) on top of the other; strictly speaking, neither can we *see* that one billiard ball is *resisting* the entrance into the space which it occupies of another ball which strikes it, but merely that the first ball is at one moment moving toward the second ball, that it then strikes it, and that the second ball then moves in a certain direction, etc. On this view, then, a common sensible is a simple idea which is an immediate

object of experience gotten by more than one sense, and solidity is not a common sensible because it can be immediately perceived only by the sense of touch.¹¹

Although I find this an interesting argument, it is not one that Berkeley is likely to have considered seriously in this context. For when Berkeley considers the question of what we immediately perceive, he is led to *his own theory* that there are no common sensibles at all, and that the proper objects of vision are signs of the proper objects of touch, specific correlations being learned through past association and particular judgments being *in all cases* the result of inference. So if Berkeley were to follow this line of argument, he would be led to his own theory, and even in terms of that theory the relation between felt and seen solidity is the very same as that between felt and seen extension or figure. But what I am attributing to Berkeley is a view which I am claiming he might have adopted had he attempted to remain within the Lockean framework. The argument under consideration, far from being an obstacle to treating all ideas of primary qualities as common sensibles within the Lockean theory, leads to just the sorts of considerations which Berkeley saw as conclusive grounds for rejecting that theory altogether.

7. Aside from the bare fact that Locke actually did or at least should have accorded the status of a common sensible to each of his ideas of primary qualities, there are even more general considerations which may have led Berkeley to adopt the thesis that I am attributing to him on the basis of the passage from the *TVV*.

Locke attempts to explain the distinction between primary and secondary qualities in a number of different ways, and I am of no mind to enter the long standing controversy as to what, if anything, the distinction really comes to.¹² I would like to point out, though, that certain of the obviously important things that Locke says about the distinction at least strongly suggest, if they do not strictly imply, that all ideas of primary qualities must be common sensibles.

On one widely accepted view,¹³ Locke's dis-

tinction between primary and secondary qualities is the distinction between what are strictly speaking the observable qualities *of the object* and what are merely powers in the object, due to its genuine qualities, to cause certain ideas *in us*. Although our ideas of the secondary qualities of an object of course depend upon (the primary qualities of) the object, they are mind dependent in a way in which our ideas of primary qualities are not. For our ideas of secondary, but not of primary, qualities depend upon the specific nature of *specific* sense receptors. The secondary qualities are thus according to Locke "nothing in the objects themselves,"¹⁴ whereas the primary qualities are "real qualities, because they really exist in . . . bodies."¹⁵ How might one defend the claim that a given idea is the idea of a primary and not of a secondary quality on this view? By an appeal to the fact that the idea does not depend upon a specific sensory receptor, but remains the same whether it is gotten by sight or touch. *In short, by finding that it is a common sensible*. And how might one defend the claim that a given idea is an idea of a secondary quality? *By finding that it is an idea proper to a particular sense*. Thus of secondary qualities and our ideas of them Locke writes:

Take away the sensation of them; let not the eyes see light or colours, nor the ears hear sounds; let the palate not taste, nor the nose smell, and all colours, tastes, odours, and sounds, *as they are such particular ideas*, vanish and cease, and are reduced to their causes, i.e. bulk, figure, and motion of parts.¹⁶

Equally suggestive is Locke's solution to the old problem of how it is that the same body of water may feel cold to one hand and hot to the other at the same time:

. . . it is impossible that the same water, if those ideas were really in it, should at the same time be both hot and cold. For, if we imagine *warmth*, as it is in our hands, to be nothing but a certain sort and degree of motion in the minute particles of our nerves or animal spirits, we may understand how it is possible that the same water may, at the same time, produce the sensations of heat in one hand and cold in the other; which yet *figure* never does, that never producing the

¹¹ This quite Humean argument was suggested to me by my colleague, Professor Howard Stein. One wonders whether Locke could have responded in this way and still said some of the things which he did about the idea of power.

¹² A recent attempt at explicating the distinction between primary and secondary qualities which at least comes close to according ideas of primary qualities something like the status of common sensibles is that of Jonathan Bennett, "Substance, Reality, and Primary Qualities," *American Philosophical Quarterly*, vol. 2 (1965), pp. 1-17.

¹³ See Reginald Jackson, "Locke's Distinction Between Primary and Secondary Qualities," *Mind*, vol. 38 (1929), pp. 56-76; W. H. F. Barnes, "Did Berkeley Misunderstand Locke?", *Mind*, vol. 49 (1940), pp. 52-57; R. I. Aaron, *John Locke* (Oxford, Clarendon Press, 1955), especially pp. 116-117.

¹⁴ *Essay*, Bk. II, ch. VIII, Sect. 14.

¹⁵ *Ibid.*, ch. 17.

¹⁶ *Ibid.*

idea of a square by one hand which has produced the idea of a globe by another.¹⁷

In the above passage, the secondary quality of warmth is contrasted with the primary quality of figure in terms of perceptual invariance solely within the realm of the sense of touch. It is but a short step, though, to extend this contrast across sense modalities, primary qualities being ones the ideas of which remain specifically the same whether received from sight or touch.

8. Whether or not Berkeley had the above sorts of considerations in mind, there is ample textual evidence, in addition to the striking passage in the *NTV*, that he accepted the principle that within the Lockean framework all ideas of primary qualities are common sensibles.

First, there is the fact that in the *NTV* Berkeley takes pains to deny, of each of Locke's primary qualities, including solidity,¹⁸ that it is a common sensible.

Secondly, and more importantly, the argument that number is not a common sensible is employed directly as a basis for denying (a) that it is a primary quality, and (b) that it really exists in things themselves. In Sect. 109 of the *NTV*, Berkeley writes that "it ought to be considered that number (however some may reckon it amongst primary qualities) is nothing fixed and settled, really existing in things themselves."¹⁹ Arguments for this conclusion, which are arguments to the effect that it is not a common sensible, are found in Sects. 109 and 110, and in Sect. 111 the general thesis, stated throughout the *NTV*, is asserted: "From what has been premised, it is plain that the objects of sight and touch make . . . two sets of ideas which are widely different from each other." In the case of number, then, Berkeley clearly moves from the conclusion that it is not a common sensible to the conclusion that it is not a primary quality and thus does not really exist in things themselves. The statement of the general thesis of the heterogeneity of the objects of sight and touch which immediately

follows in Sect. 111 leads me to believe that he would have been willing explicitly to make the same sort of move with respect to Locke's other primary qualities. Be that as it may, it is clear that the inference which Berkeley clearly made with respect to number is valid only if he is granted the principle that any idea which is not a common sensible is not an idea of a primary quality. But this is of course logically equivalent to the thesis that I am attributing to him; namely, the principle that all ideas of primary qualities are common sensibles.

9. If I am correct that Berkeley believed he had destroyed the distinction between primary and secondary qualities by demonstrating in the *NTV* that there were no common sensibles, the question naturally arises as to why he did not appeal to this argument for the same purpose in the *Principles* and the *Three Dialogues*. The answer, I believe, is that he thought he had found a more general argument which would permit him to deny the existence of both common sensibles and primary qualities without bothering to relate them explicitly. That argument is the argument against abstract ideas which appears in fully developed form in the lengthy Introduction to the *Principles*.

That Berkeley believed that all common sensibles were abstract ideas is made just as clear in the *NTV* as it is made clear in the *Principles* that he believed that all ideas of primary qualities were abstract. In Sects. 122-125, 127, and 130 of the *NTV* arguments to the effect that it is impossible to form abstract ideas of extension, figure, and motion are presented as independent grounds for denying that they are common sensibles. Of extension, for instance, Berkeley writes, in Sect. 122:

I am apt to think that when men speak of extension as being an idea common to two senses, it is with a secret supposition that we can single out extension from all other tangible and visible qualities, and form thereof an abstract idea, which idea they will have common both to sight and touch.

¹⁷ *Ibid.*, ch. 8, Sect. 21.

¹⁸ See Sects. 45 and 135 of the *NTV*.

¹⁹ Turbayne, in his edition of Berkeley's *Works on Vision* (Indianapolis, Bobbs-Merrill Company, 1963), footnotes the word "some" in this passage and calls attention in the note to the passage from Aristotle quoted by me in note 9.

Could it be that Turbayne is simply assuming the correctness of the argument in this paper? In light of the difficulties which I have dealt with in Sects. 4-6, this would certainly be to assume quite a lot. Even more perplexing is D. W. Hamlyn's remark on p. 103 of his *Sensation and Perception*, (London, Routledge and Kegan Paul, 1961), that

Having made the distinction between primary and secondary qualities he [Locke] had to admit that one feature of the former is that they are perceptible by more than one sense. Berkeley's denial of the distinction entailed the denial of the consequence.

Although Hamlyn would, so to speak, have Berkeley working in the opposite direction from the one I have argued he may have worked in, Hamlyn provides no evidence whatsoever that Berkeley made any connection between the two distinctions.

And immediately follows, in Sect. 123, a refrain familiar from the *Principles*:

Now I do not find that I can perceive, imagine, or anywise frame in my mind such an abstract idea as is here spoken of. A line or surface which is neither black, nor white, nor blue, nor yellow, etc., nor long, nor short, nor rough, nor smooth, nor square, nor round, etc., is perfectly incomprehensible.

10. If one grants my thesis that Berkeley believed that all ideas of primary qualities were to be accorded the status of common sensibles within the Lockean framework, the fact that he believed that both common sensibles and ideas of primary qualities were abstract ideas suggests an even broader reconstruction of what he could have had in mind from the writing of the *Philosophical Commentaries* onward.²⁰ This reconstruction, in outline, is as follows: The immaterialism, as well as the theory of vision, was well developed in the *Philosophical Commentaries*. Realizing that the theory

of vision contained an answer to an obvious objection to his immaterialism, and having abandoned the plan (revealed in the *Philosophical Commentaries*) of encompassing both the theory of vision and the immaterialism in a single work, he turned first to the better developed theory of vision. Although the *NTV* speaks (with the vulgar) as if tangible objects do really exist without the mind, Berkeley believed that the destruction of the doctrine of common sensibles carried out in the *NTV* was tantamount to the destruction of the distinction between primary and secondary qualities. In the course of developing his argument against the doctrine of common sensibles, Berkeley developed the general argument against abstract ideas. Having all along believed that ideas of primary qualities were common sensibles, he lost no time in using the argument against abstract ideas as an argument against primary qualities, publishing it as the Introduction to the *Principles* one short year after the publication of the *NTV*.²¹

University of Minnesota

Received August 12, 1968

²⁰ Berkeley may have gotten started on the train of thought which led both to his theory of vision and his criticisms of Locke's distinction between primary and secondary qualities through a consideration of the problem raised by Molyneux about the man born blind being made to see, discussed by Locke in Book. II, ch. 9, Sects. 8-9 of the *Essay*.

²¹ I have profited greatly by discussing an earlier draft of this paper with Professors Chin-Tai Kim and Howard Stein.

VIII. MORALITY AND THE PARADIGMATIC INDIVIDUALS

ANTONIO S. CUA

IN *The Language of Morals*, R. M. Hare maintains that it is in practice impossible to specify completely a way of life of which a person's decisions of principle form a part. We are told that "the nearest attempts are those given by the great religions, especially those which can point to historical persons who carried out the way of life in practice."¹ This emphasis on the importance of historical persons gains significance in a man's moral life only when they are in some manner viewed as living exemplars of conduct.² Thus it has been recently suggested by Kai Nielsen that "for the most part, people get their standards not from ethical treatises or even scriptural texts or homely sayings, but by idealizing and following the examples of some living person or persons."³ If this suggestion is sound, it points to the view that in some fashion certain persons are a source of actual conduct from the point of view of a moral agent. The best method of reflection appears to be a study of at least some founders of living religions to determine in what way they enter into the lives of moral agents in various moral practices. This interesting study has been carried out in Karl Jaspers' exposition of Socrates, Siddharta Gautama, Confucius, and Jesus as four paradigmatic individuals who have "exerted a historical influence of incomparable scope and depth." Jaspers states that "no single type can account for these four," for "their historicity and consequent uniqueness can be perceived only within the all embracing historicity of humanity, which each of them expresses itself in a wholly different way."⁴ Moreover, instead of explicating the notion of the paradigmatic individuals, Jaspers directs his work

primarily to doctrinal exposition and exegesis. In this essay I wish to reflect on this notion and to suggest how it may be of interest to moral philosophy.

I

Jaspers notes that Socrates, Buddha, Confucius, and Jesus "became models for mankind without setting themselves up as examples."⁵ Each of them appears as a unit class, but also as a model that is exemplary in some way. An exemplar is not literally an example, otherwise it would be an example of a type. It is more like a Goethean *genuine symbol*—"a particular instance which is coalescent with a universal and which thereby plays a unique role by revealing, in a way which no particular could quite do, the nature of that more general something."⁶ And yet it cannot be a particular instance of a type, but a particular that functions in some universal or paradigmatic way. The term "model" as applied to these individuals is also misleading, for it suggests something to imitate. Although it is possible to imitate the conduct of these historical persons, imitation does not necessarily portray the spirit which appears to be embodied in the lives of these persons.

How then can a particular person function as a paradigmatic individual?

Every living person is placed in a *historic situation*. He lives within the orbit of his cultural inheritance in the face of a changing environment. In the midst of a moral crisis, he must decide whether his

¹ R. M. Hare, *The Language of Morals* (Oxford, 1952), p. 69.

² See "The Logic of Confucian Dialogues," in *Studies in Philosophy and the History of Philosophy*, vol. 4 (1969), ed. by John K. Ryan, where I have attempted to exhibit the role of historical personages as paradigmatic individuals in Confucian ethics.

³ Kai Nielsen, "Why Should I Be Moral?", *Methods*, vol. 15 (1963). Reprinted in Paul Taylor's *Problems of Moral Philosophy* (Belmont, 1967), p. 516.

⁴ Karl Jaspers, *Socrates, Buddha, Confucius, and Jesus: The Four Paradigmatic Individuals*, tr. by Ralph Manheim (New York, 1957), p. 3.

⁵ *Ibid.*, p. 95.

⁶ Quoted by Philip Wheelwright, *Heraclitus* (New York, 1964), p. 14.

encounter with the cultural tradition inherent in his moral practice will take the form of rejection or critical acceptance. To reject the tradition *in toto* is to cut oneself off one's historic roots. It is in a sense to lose one's personhood, for one loses his anchor of identification. The use of language is by itself an indication of our historic roots. Ludwig Wittgenstein says, "to imagine a language means to imagine a form of life."⁷ Speech is "part of an activity, or of a form of life."⁸ Language, as a form of life, embodies a tradition that furnishes the source from which our activities may gain a practical import. Moral discourse, in particular, reflects certain pre-established standards and criteria for assessing conduct. To claim an absolute separation from the tradition without the use of language is to remain silent. Socrates, Buddha, Confucius, and Jesus undoubtedly have their moments of silence, of reflection of their own predicaments. But they do speak against the backgrounds of their situations. In so doing, however partial, they must acknowledge their historic reality.

Speech, moreover, is not an indication of mere acceptance, for each manner of speaking is a *style of orientation* to a tradition. It may be an act of critical acceptance as in the case of Socrates in *Crito*, or of an act of critical transformation as in the case of Confucius. It may also be an act of acceptance to achieve transcendence from existential predicaments as in Jesus and Buddha. The uniqueness of these men appears to lie in their peculiar *styles*.

In his manner of relating himself to a tradition, the paradigmatic individual in effect transfigures the content of the tradition into his own life. The effect is one of transfiguration. Herein lies the supreme venture—the risk of disapproval and condemnation of his fellow beings. This is perhaps the reason why the paradigmatic individuals can both inspire devotion and enthusiasm, and at the same time evoke hostile response from their audience. Transfiguration can be achieved only after a struggle. Once it is achieved, its novelty and suasive power become accepted by the majority. It may become the source of a new

tradition, but in itself, as a living person, it is not a tradition. This may account for the fact that the lives of the paradigmatic individuals cannot really be our own. Jaspers remarks that "in general, the content of their thinking cannot be ours, but their manner of thinking can show us the way."⁹ However, the paradigmatic individual may be a source upon which my life may build, but my life may also crumble in so doing, for the source is hollow, the style cannot be my own. As Alcibiades said of Socrates, "I don't know whether anybody else has ever opened him up when he's being serious, and seen the little images inside, but I saw them once, and they looked so godlike, so golden, so beautifully and so utterly amazing that there was nothing for it but to do exactly what he told me."¹⁰ What makes these images significant is precisely the *transhistorical possibilities* that seem to be suggested by the paradigm. Although the paradigm is situated in a historic present, it seems to project itself in a prospective historical future. In this sense, each may be viewed as a Leibnizean monad pregnant with future possibilities. These future possibilities, to be significant to any person, have to be individuated. In this individuating concretizing process, the paradigm may function, on the one hand, as a source of retrospective justification for moral agents, and on the other hand, as a prospective task to be accomplished.¹¹ From the retrospective point of view, the paradigmatic individual is an embodiment of a set of moral principles and rules that an agent may appeal to for justification of his acts. From the prospective point of view, the paradigmatic individual is an embodiment of moral ideals which sets the task for personal moral accomplishment. Since the paradigmatic individuals like Confucius, Jesus, Buddha, and Socrates are a source of morality, living in the knowledge of this source and living in accordance with a tradition is an achievement.

We may note that with the possible exception of Jesus, none of the individuals mentioned claimed himself to be paradigmatic. Socrates was a man in earnest quest for self-knowledge that is compatible with his view of man's characteristic rationality.

⁷ Ludwig Wittgenstein, *Philosophical Investigations* (New York, 1953), Pt. I, §19.

⁸ *Ibid.*, Pt. I, §23.

⁹ Jaspers, *op. cit.*, p. 96.

¹⁰ *Symposium*, 216e–217a. Translated by Michael Joyce in E. Hamilton and H. Cairns (eds.), *The Collected Dialogues of Plato* (New York, 1961). Cf. Lao Tzu, ch. 70; "The Sage wears coarse clothes while keeping the jade in his bosom," Lao Tzu: *Tao Teh Ching*, tr. by J. C. H. Wu (New York, 1961).

¹¹ The distinction between the retrospective and prospective views of morality was developed in my "Toward An Ethics of Moral Agents," *Philosophy and Phenomenological Research*, vol. 28 (1967), pp. 163–174.

Confucius was a man who hoped to reinstitute an ideal of a harmonious social order which he thought to be implicit in the Chou dynasty. Buddha bowed in silence regarding his own liberation from the wheel of *samsara*. Socrates did not proclaim the success of his quest. In accordance with our knowledge drawn from *The Analects*, Confucius never asserted himself as being in any way paradigmatic. Each of the four men lived his life as a moral pioneer within an established tradition. The success of their pioneering tasks is a testimony to their moral genius.

II

How do the four paradigmatic individuals acquire their historic significance and moral relevance? All four were moral pioneers or reformers. From the point of view of the psychology of personality, they are in Max Weber's sense *charismatic* persons, possessing, in the eyes of their followers, a set of unusually attractive qualities or powers of moral leadership; a sense of devotion to a mission and confidence in its rightness; and also an ability to display these qualities in concrete situations of doing and suffering. A charisma is "always alien to the world of everyday routine: it calls for new ways of life and thought."¹² This accounts for their innovative and revolutionary character. From the point of view of the content of their teachings, their utterances have two prominent characteristics. First, the moral principles and ideals which they espoused and preached are vague, for the principles and ideals announced do not completely specify the rules or types of acts falling within a well-defined scope of application, nor do they spell out the criteria for determining the appropriate situations to which the principles are intended to apply. The room is thus left for individual interpretation and decision. Secondly, the same utterances adjudged by posterity to be vague appear to have their suggestive and persuasive relevance to concrete situations of human concern. We may here observe that the vagueness of their sayings owes much to the viewpoint of posterity. To the disciples of Jesus, Confucius, Buddha, or Socrates, the masters' remarks might have appeared to have a crystal clarity. Their spoken words had perhaps an imperative and directive force. Their clarity owed their timely and felicitous relevance to actual situations at the

time the utterances were made. To their immediate audience, the spoken words seemed to solve the hearers' problems. The success of these remarks terminates as soon as the problematic situations become unproblematic ones. At the same time these solutions acquire a universal function. The insight of these moral pioneers lies in their reflective view of the human predicament to be met with by every moral agent. Since the predicament is a universal one, insightful solutions that liberate moral agents from the predicament will acquire relevance in moral conduct. The forceful relevance of the utterances of the paradigmatic individuals bears on their capacity to relate themselves to existential moral crises by way of providing a new style or way of life in which a living person may organize his own ideals and aspirations. The insights may then become universalized rules for later generations. But for any paradigmatic insight to retain its dynamic and persuasive force, it cannot be incapsulated in a book of rules and precepts to be followed and obeyed, but as a set of guidelines that acquires regulative authority through the practical test of its capacity to bear continuing relevance to concrete problems of deliberation and choice. In the words of Aristotle, "in the field of moral action truth is judged by the actual facts of life, for it is in them that the decisive element lies."¹³

It is one thing to express appreciation for the power of a paradigmatic individual and another to view him as a regulative source of a moral practice. Appreciation of a paradigm may impress us of its suasive powers and personal qualities that command respect and admiration, but the application of the paradigm's utterances compels us to recognize the paradigm as a source of moral practice, as generative of particular solutions to particular problems. Viewed apart from their original homes, the paradigms become universal symbols of striving for Buddhistic Nirvana, Christian union with God, Socratic self-illumination, and Confucian cosmic harmony. With a little less detachment from their historic loci, the symbols are transformed into images of insight into particular living situations. As symbols the paradigms tend to float in the realm of abstraction or ideality. As images they find the concrete locus that gives them life.

In a sense the continuing relevance of a para-

¹² Robert C. Tucker, "The Theory of Charismatic Leadership," *Daedalus*, vol. 99 (1968), p. 737.

¹³ Aristotle, *Nicomachean Ethics* (Indianapolis, 1962), 1179a, pp. 15-20.

paradigmatic individual in the life of a moral agent is a progressive articulation of its relevance to the agent's problem of choice and decision. When the moral principles learned from paradigmatic persons are assented to and adopted by a moral agent, these principles may then function as promissory notes to be reckoned with, leaving, however, the terms of fulfillment to the appropriate circumstance of individual judgment and decision. In this way a moral agent can also be paradigmatic in a more modest way, in the manner in which he exercises his judgment of application of principles. The judgments are in effect his rulings on the relevance of the paradigms, even though by his actions he may not succeed in a scope of influence comparable to that of the four paradigmatic individuals. His judgments and corresponding actions may have only a limited scope of universality, and consequently, a restricted scope of relevance to other persons.

III

What then accounts for the universality of function in the paradigmatic individuals?

The difficulty of answering this question lies in the problem of description of the paradigmatic individuals. If we grant that each of the four is *unique* in some way, their uniqueness seems to be an elusive thing. It seems to escape any attempt at an adequate description. For any description imports criteria that are foreign to the life of the person. In addition to this difficulty, in relating the lives of these paradigms, our description cannot be easily distinguished from our evaluation. As Geoffrey Warnock recently pointed out, unlike legal proceedings, "ordinary discourse is not, since it does not need to be, similarly regimented." If I am relating to someone, for instance, "about the career of Mussolini, it would be unrealistic to look for—to assume that there must be—a point at which description of his doings terminates, and evaluation of them begins."¹⁴ This is, of course, not a claim that the distinction between description and evaluation cannot be made, but a claim that in ordinary discourse "one might expect to find description and evaluation so inextricably intermingled as to constitute, as it were, a seamless garment."¹⁵

We may here venture a proposal in terms of the distinction between the paradigmatic individuals

and ordinary men. We may say that what distinguishes the former as paradigms lies in the fact that they *generate* their own criteria for evaluation. Their uniqueness lies in their *styles of life* rather than in their content. The language of criteria has, so to speak, a retrospective significance from the point of view of a moral agent. Bits and pieces of speech and conduct thus function as criteria only *ex post facto*, and then only to other men.

The notion of "generation of criteria" is more than just a metaphor. It is used to capture the living significance of the paradigmatic individuals. The term is used to focus on the *creativity* of a moral life that prospectively functions as an embodied set of criteria for other people, and retrospectively as a source of justification. Both the prospective and the retrospective uses of paradigms are an attempt at a sort of description from the point of view of their regulative significance to a moral agent. As a set of criteria for evaluating conduct, the paradigms seem to exemplify them in their lives, but such exemplifications amount to a way of speaking that stresses on the *universality* of function of a particular person. As a particular *qua* particular, the criteria viewed by moral agents as implicit in the paradigms are in reality extrinsic to the lives of these paradigmatic individuals. In other words, a particular person acquires paradigmatic authority by virtue of its cognitive and persuasive relevance to the living problems of the moral agents. Of itself a particular individual cannot endow with any effectiveness its own universality, for there are, strictly speaking, no self-authenticating acts of any individual. A self-appointed paradigm can at most be a possibility of a paradigm. Being a paradigm is an appreciation and recognition by others. It is a recognition of a living spirit in a moral history. No self-authenticating act is ultimately possible, from the point of view of rational persuasion, without a backing from an external source, whether that of divine revelation, of established historical authority, or of political and social sanctions. But none of these can have any force without an act of ascription by others.

Thus an individual, without his authority being assented to by others, cannot pronounce his own utterances as authoritative. He has to present legitimate credentials for his authority to be recognized. Uniqueness of moral creativity is a

¹⁴ Geoffrey Warnock, *Contemporary Moral Philosophy* (London and New York, 1967), p. 64.

¹⁵ *Loc. cit.*

style that is at the same time an achievement. The uniqueness of a style of life is not a given but an accomplished fact. If the notion of a moral agent is to be construed as a unique person that admits, according to Kant, of neither market nor fancy price, it is because the dignity of persons is an award, not an intrinsic characteristic of human beings *qua* human beings. Since we do not award uniqueness to everyone, we must have some grounds for doing so. Paradoxically, the uniqueness of the paradigmatic individuals consists by and large in the criteria that they project and sometimes articulate for their followers and posterity, for they generate their own criteria. They bespeak their own norms of conduct that serve as guides for others. From our own point of view as moral agents, the criteria must function consistently in their various and varying applications. But this consistency, as it has been recognized in recent moral philosophy, is a requirement of rationality, but not a requirement of a paradigmatic life. The larger sense of consistency is at stake here—a consistency that is the fitting and the appropriate. It is the consistency of a rhythm of a life, of an on-going process of conduct throughout a man's life, not that of one proposition with another.

The claim that I am different from other men is not an announcement of my uniqueness of being a person. Descriptively, the focal difference lies in the bio-cultural space that is occupied by a man as a living organism. To point to this difference is to point to a distinction that need not be a mark of *distinction*. The ambiguity of the notion of distinction moves between the descriptive and the normative. Uniqueness is a difference that is normatively distinctive. It is a normative notion of merit. Being uniquely paradigmatic as a person is an endowment or entitlement given by others by virtue of the fact that the person in question possesses certain qualities judged to be meritorious on certain acknowledged criteria. Although a person cannot authenticate his utterances or actions with any effectiveness, he can present grounds for such an authentication by others. He can present merit-claims, but these claims still have to be assessed before an acknowledgement of their authority can be given. The notion of the dignity and worth of a person is a dubious notion if it is intended as a way of asserting the intrinsic property of being a person. Moreover, the use of this notion has a persuasive force, for it is an effective weapon in a democratic society against the oppression of

individual liberty. Neither the belief nor the exercise of liberty assures us of any recognized merit as a person, but it does point to the importance of social provision of opportunities for individuals to acquire merits.

IV

The four paradigmatic individuals are the living foundations of personal moralities. They provide the source of principles and ideals, and also the support for certain continuing moral practices. The uniqueness of their styles, their manner of organizing their lives, have been an inspirational source of moral guidance. This at any rate has been true of the majority of moral agents in these moral practices. The significance of the notion of paradigmatic individuals thus lies in its focus upon a foundation of personal morality in terms of its historic and dynamic source rather than in terms of an abstract model of reason and sentiment. Like the paradigmatic individuals, the moral agent also lives within a given moral tradition. He must exercise his judgment concerning the significance of the ideals and principles implicit in his moral practice. His *practical wisdom* consists in his ability to mold the existing tradition into his own style of life—an organized pattern of living that gives meaning to his aspirations and values. From his point of view, a paradigmatic individual stands as a preeminent exemplar of excellence and virtue, for the paradigm provides a source of both moral principles and values. The rules and principles that a paradigm embodied are an agent's articulation of the principal features of a life with a style of its own. In the same sort of way, the moral agent carries within himself the burden of a tradition—a continuing burden to make the moral practice, originally inspired and founded by the paradigm, relevant to *his* problems and to the common predicament of humanity. The *spirit* of the paradigm is the spirit of moral creativity within a moral practice. A historic source of a moral practice, in this way, gives room also to emergent personal novelties.

In the paradigmatic individuals, reason and sentiment are not distinct, words and actions do not form an articulate inference. Principles and actions are not placed in a logical map. The paradigm, detached from its life, may serve as a map with a set of principles of morality that functions as vessels to be filled rather than substance to be ingested. Since their words and actions

are responses to the problematic situations which they encountered, these words and actions cannot literally be our own. Their particular insights and seeming appropriateness of conduct are tied to their historicity. For us, as moral agents, we still need to locate the fittingness of these insights within our own psychological, social, and cultural milieu. Our creativity thus consists in our ability to give the principles implicit in our moral practice a substance and meaningful content. We also act in accordance with them with a style of our own. We may not be paradigmatic in the same sense as were Socrates, Confucius, Jesus, and Buddha. But some of our actions when fitting and appropriate to living situations may be paradigmatic for others, for our children, our friends, and acquaintances. And in extending the scope of our activities and concern for others, some of our actions may be paradigmatic even for persons who may not have any special relation to us. This is perhaps the point of Chu Hsi's advice: "In one's words there should be something to teach others," and "in one's activities there should be something to serve as a model for others."¹⁶ As a source of value our paradigm is an ideal to which we aspire. In acting

in consonance with an ideal we may also generate ideals for others.

Thus from the point of view of a moral agent, behind his living moral practice or tradition, there are paradigmatic individuals that form its inspiration and sustain its life. However, since the content of these individuals' lives cannot be our own, there is still room for moral creativity of agents within a moral practice. As Jaspers succinctly remarks, the paradigmatic individuals are the "beacons by which to gain an orientation, not models to imitate."¹⁷

The topic of moral creativity within a moral practice does not occupy a central place in recent ethical discussion. The problem, as I understand it, is not one of providing decision procedures nor of finding an adequate theory of moral judgments without a prior examination of the structure of moral practice. It is a problem of exploring the structure of moral decision in terms of the various dimensions that appropriately play their roles in both personal judgment and conduct. Now is perhaps the time to give the notion of moral pioneer a more than an incidental role in moral philosophy.

Catholic University of America

Received October 28, 1968

¹⁶ W. T. Chan (tr.), *Reflections on Things at Hand: The Neo-Confucian Anthology Compiled by Chu Hsi and Lu Tsu-Ch'ien* (New York, 1967), p. 76.

¹⁷ Jaspers, *op. cit.*, p. 96.

IX. THE ROLE OF LANGUAGE IN PHENOMENOLOGICAL ANALYSIS

GUIDO KÜNG

THE main difficulty which analytic philosophers have with the phenomenological method centers around the notion of an "eidetic intuition" (*Wesensschau*). To meet them head on it would be necessary to explain in what sense the term "eidetic intuition" has to be understood and to show that such intuitions do occur. However, in this paper I shall make only a modest preliminary contribution to this great task. By clarifying the important role which language is rightfully playing in phenomenological analysis, I hope to remove a first stumbling block which hinders an encounter between phenomenologists and analytic philosophers.

There can be no doubt that the phenomenologists are strongly interested in language as an *object* of investigation. Remember that Husserl had a special reason for being interested in the words of spoken or written language: he had to overcome the then reigning subjectivist psychologism and to defend the objectivity of the content of our thoughts; and this objectivity comes to the fore the most clearly when this content is considered as attached to the expressions of a language which is intersubjectively understood. In his "*Logische Untersuchungen*" Husserl was therefore concerned with extremely detailed linguistic and semantical investigations.

However, the *methodological* question concerning language as a *tool* of analysis has been rather neglected in the phenomenological literature. This is unfortunate since for the analytic philosophers this question is of primary importance. But many phenomenologists are not interested in discussions with analytic philosophers. They consider them as "nominalists" who talk about words, when what

is important are "the things." The slogan of the 14th century realists: "*Nos imus ad res, de terminis non curamus*"¹ is still very much *en vogue*.

This does not mean that phenomenology has nothing to say about the role of language as a tool of analysis. In Husserl's *Ideen*, for instance, we find a brief section on the *Faithful Expression of the Clearly Given* and a note *On Terminology*.² But I know of only one exposition of the phenomenological method which considers the role of language in some detail. It has been published by Roman Ingarden in 1919-1920 in a Polish journal.³ Since Ingarden's statements are thus unique, in what follows I shall often quote them *in extenso*.

I

Phenomenology is in many respects like ordinary language philosophy. It is dissatisfied with the conceptual frameworks of previous philosophies and objects that the philosophical questions have been articulated in misleading ways. As a cure it prescribes a return to the "things themselves," to an original experience, unprejudiced by any speculative thinking. It strives for an experience which would not yet be clothed in the concepts of any theory. But, of course, there is no completely "naked" experience. It is impossible to get rid of *all* previous concepts, "as if we could somehow prise the language off the world."⁴ Ingarden admits this explicitly: "It is not possible in practice to stop one day to use all concepts and to limit oneself exclusively to an 'immediate intuition' of the objects."⁵ At the very moment when we start a phenomenological investigation, he says, "we already possess everyday language, of which we

¹ F. Ehrle, *Der Sentenzenkommentar Peters von Candia* (Münster, 1925), p. 322.

² E. Husserl, *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie I*, Husserliana, vol. 3 (The Hague, 1950), pp. 154-155 and pp. 206-207.

³ R. Ingarden, "Dążenia fenomenologów" (The aims of the phenomenologists), *Przegląd Filozoficzny*, vol. 22 (1919-1920); reprinted in *Zbadanie nad filozofią współczesną* (Investigations on contemporary philosophy), (Warsaw, 1963), pp. 267-379.

⁴ Charles Taylor, "Phenomenology and Linguistic Analysis," *Proceedings of the Aristotelian Society*, Supplementary vol. XXXIII, (1959), p. 102.

⁵ Ingarden, *op. cit.*, p. 302.

make use to settle (and usually with success) the diverse affairs of daily life. This ordinary language contains—both in its vocabulary and in its different syntactical functions—many ordinarily established concepts which concern, among others, the objects which we want to analyze in immediate experience.”⁶ (Notice that at the time when this was written, ordinary language philosophy—as we know it today—did not yet exist.)

II

But exactly “What do we need ordinary language for at this juncture?” Ingarden answers his question:

We need it basically—though not always—in order to be able to *choose* which object, or which class of objects, we want to investigate more closely. . . . The concepts especially the general ones, have the property of selecting from the whole of all the objects of our environment a certain set of objects, on which we can focus our attention.”⁷

Also we usually need to have words, and not merely disembodied concepts, because the focusing of our attention should be relatively stable:

It is very useful to have a stable symbol as a supporting point of our intention, because it facilitates the task of keeping the direction of our intention fixed (within certain limits).”⁸

In Husserl’s terminology such a focused attention is called a meaning-intention (*Bedeutungsintention*) or a signitive or symbolic intention.⁹ Knowledge (*Erkenntnis*) is the fulfillment of such a meaning-intention. Husserl recognizes that one can have no knowledge without having a meaning-intention. However, he does not discuss explicitly how much

the range of our present meaning-intentions is limited because of our language.

But that Husserl did not explicitly investigate how our experience is preconditioned from a linguistic point of view must not be interpreted as if he had been unaware of the fact that our present experience is preconditioned by our past experience. Quite the opposite is true. He described in very much detail what he called the “horizon” of perception, namely the fact that every single perception necessarily refers beyond itself, i.e., necessarily fits into a larger horizon. This means that our present perception gets much of its meaning from our previous perceptions. To use the technical terminology of phenomenology: what we perceive as given now, the present *noema*, is constituted not merely on the basis of the present *hyle*, the “sensations” (*Empfindungen*) now present in the stream of consciousness, but also on the basis of our previous experiences.¹⁰

Van Peursen, e.g., has noticed the analogy between this horizon of perception and a linguistic “horizon” and has compared what he calls Husserl’s “rules” of perception with Wittgenstein’s rules of language games.¹¹

III

However, our meaning-intention may become only “partly” fulfilled or not fulfilled at all,¹² and then it has to be changed or at least improved. This leads us to another important point concerning ordinary language.

For the phenomenologist there is no guarantee that the concepts of ordinary language are completely adequate. We have to be critical even with respect to ordinary language. An absolute reliance

⁶ *Ibid.*, pp. 305–306. Cf. how Husserl, *op. cit.*, states also that phenomenology uses “ordinary language” (*die allgemeine Sprache, der gemeine Wortgebrauch*), but that ordinary usage is often vague and has to be improved.

⁷ Ingarden, *op. cit.*, p. 306.

⁸ *Ibid.*, p. 313.

⁹ Cf. E. Husserl, *Logische Untersuchungen*, vol. 2, pt. 2, 3rd ed. (Halle, 1922), pp. 32f.; J. N. Mohanty, *Edmund Husserl’s Theory of Meaning* (The Hague, 1964), pp. 36f.

¹⁰ Cf. Husserl, *Ideen I*, pp. 199–203, pp. 366–368. For Husserl’s doctrine concerning the “sensations” (*Empfindungen*) cf., e.g., *Logische Untersuchungen* vol. 2, pt. 1, pp. 381–397, esp. p. 274. It has to be emphasized that this doctrine implies no “myth of the given” in the empiricist sense: it accepts no sense-data as ultimate objects of perception. Husserl’s “*Empfindungen*,” which he describes as being the ultimate *hyle* (matter) on which our knowledge is based are not *objects of perception*, but are only “*appereived*,” “*lived through*” (*erlebt*) (*op. cit.*, p. 385). All phenomenologists stress that what is “given” in ordinary perception is an object which, *qua* object, has been constituted by the perceiving subject. A. Gurwitsch even maintains that pure sensations cannot be reached in phenomenological analysis (Cf. his book *The Field of Consciousness*, [Pittsburgh, 1964], pp. 92f.). Their status in phenomenology is very similar to that of the “inner episodes” which W. Sellars postulates as a kind of theoretical entities (cf. his book *Science, Perception and Reality* [London, 1966], pp. 181f.).

¹¹ C. A. Van Peursen, “Edmund Husserl und Ludwig Wittgenstein,” *Philosophy and Phenomenological Research*, vol. 20 (1959), pp. 181–197.

¹² Husserl, *Logische Untersuchungen*, vol. 2, pp. 41f.

on the conceptual framework of our native language may make us blind for important features of the "things themselves." Ingarden writes:

Once we have come into direct contact with the objects, the role of the concepts taken from ordinary language comes to an end; we must go on and strive to make ourselves independent from all the suggestions which flow from this language. For there is no doubt that the habit to use just such words, resp. concepts, and not other ones, produces (or at least may produce) a certain bias in the way in which we see the objects (*idola fori*).¹³

However, here again the phenomenologist does not require that we forget suddenly all our concepts. This would be absurd since philosophers cannot become "speechless" infants again. What he asks for is rather a kind of *epoche*. According to Ingarden "we must suspend consciously the validity both of the concepts which we have been using up to now and of the judgments which we have been entertaining."¹⁴ Husserl states that "in the beginnings of phenomenology all concepts or terms must in a certain sense remain fluid."¹⁵ Instead of keeping our concepts rigidly fixed, we must be ready to modify, differentiate, and correct them throughout our phenomenological analysis.

If a description does not "fit" the phenomena, ordinary language philosophers would normally claim that we have not been using our language properly, that we have to try to give a more accurate description by making a better and more careful use of the resources of ordinary language. Phenomenologists, however, go farther than that and claim that the meanings of many words of ordinary language are inadequate and have to be changed.¹⁶ In this they are similar to the logistic philosophers who aim at the "explication" of ordinary concepts.¹⁷

However, as we shall see in more detail below,¹⁸ the new meanings have to be explained first, or at least hinted at, by means of words which still have their "old" meaning. E.g., Charles Taylor was right when he said:

We can only send in our concepts for overhaul as it were, in small groups. To "suspend" one concept for reexamination requires that others are taken for granted to carry out this examination. Our entire conceptual machinery cannot be repaired in one fell swoop . . .¹⁹

But the important point is that phenomenologists think that the "conceptual machinery" of ordinary language can and must be "repaired," that they proceed to the development of a new technical vocabulary.

IV

This implies of course the conviction that we can somehow transcend the limits of our present language. It is interesting to see that it seems that this means not only that we are able to go from our present language to an improved future language, but also that already in the present our understanding goes beyond what is expressed in our present words and formulations.²⁰ This becomes apparent, e.g., when I read a translation which has been made of a paper of mine. Even if the translator is very competent I always find passages for which I myself can give a better translation, i.e., a translation which fits more exactly what I have had "in the back of my mind" when I wrote the paper. One might be tempted to say that this is due to my shortcomings, that in my paper I did not express precisely what I had in mind. The point is, however, that it seems to be impossible in principle to fully express what one has in mind.²¹ Those familiar with mathematical logic might here be reminded of the theorem of Kurt Gödel according to which no formal axiom system can express all that mathematicians have "in the back of their mind" when they think of the arithmetics of natural numbers.

Philosophers of the "continental" phenomenological-existentialist tradition express this point often in obscure and esoteric sayings, which philosophers used to clear and plain English cannot understand. For instance Professor Gadamer

¹³ Ingarden, *op. cit.*, p. 306.

¹⁴ *Ibid.*, p. 307.

¹⁵ Husserl, *Ideen I*, p. 206.

¹⁶ Cf. Husserl, *Logische Untersuchungen*, vol. 2, pt. 1, p. 30 where the author introduces the term *Ausdruck* as a technical term with a meaning that differs from the one which it has in "ordinary language" (*in normaler Rede*), and where he stresses that we have often "to force language" (*der Sprache Zwang antun*).—Actually, e.g., Austin admits explicitly: "In spite of the wider and acute observation of the phenomena of action embodied in ordinary speech, modern scientists have been able, it seems to me, to reveal its inadequacy at numerous points . . ." and "certainly . . . ordinary language is *not* the last word: in principle it can everywhere be supplemented and improved upon and superseded. Only remember, it is the *first* word" (cf. "A Plea for Excuses" *Proceedings of the Aristotelian Society*, vol. 57 (1956–1957), p. 11 and p. 29). Of course, at the point where one goes beyond ordinary language, one is no longer doing ordinary language philosophy.

¹⁷ But see also Sect. VI below.

¹⁸ See Sect. VII.

¹⁹ Taylor, *op. cit.*, p. 103.

²⁰ This point is not mentioned by Ingarden.

²¹ I owe this example to Professor N. Luyten of Fribourg.

writes: "... die endlichen Möglichkeiten des Wortes (sind), dem Sinn wie einer Richtung ins Unendliche zugeordnet . . .," i.e., "the finite possibilities of the word are correlated with the sense as with a direction into the infinite" and "In der Aussage wird der Sinnhorizont dessen, was eigentlich zu sagen ist, mit methodischer Exaktheit verdeckt," i.e., "In the proposition the meaning horizon of what is actually to be said is masked by methodical preciseness."²²

V

But how can we get a better grasp of what we have "before our eyes" or even already "in the back of our mind"? The uninitiated imagines often that phenomenology simply recommends to "wait and see" i.e., to gaze passively at the things and to wait for a mysterious illumination. Actually we get more valuable advice than this. The key word is "free imaginative variation."²³ In a way which is similar to the procedure of the geometrician who surveys in his imagination all possible variations of a triangle, the phenomenologist tries to survey the possible variations of the phenomenon under analysis. We should start with a few concrete situations which are clear instances of the phenomenon we want to investigate (e.g., to take a simple example, "Mrs. Stone giving an apple to her child" is a clear instance of the phenomenon of "giving"). Then we systematically imagine slight changes, variations, of these situations (e.g., "Mr. Smith from the grocery store gives an apple to the child," "Mr. Smith from the grocery store gives an apple to the child and gets a dime in return"). If a change strikes us as giving an entirely new character to the situation, then we have discovered an essential feature of the phenomenon (e.g., the case where Mr. Smith gets paid for the apple is significantly different; we have no longer the previous kind of giving which was essentially gratuitous).

The phenomenologists did not fully appreciate how important a role the sensitivity for the proper use of ordinary language is playing in this procedure. Above we quoted a typical affirmation of Ingarden which stated that once the contact

with the things themselves was established, the role of language had ended.²⁴ But today the similarity between the method of free imaginative variations and Austin's "linguistic phenomenology" has been pointed out, e.g., by A. J. Ayer.²⁵ The point is simply the following: an important clue that a new situation has an essentially new character is often given by the fact that a word, or a phrase, which properly characterized the situation imagined before, sounds odd when applied to the new situation (e.g., in the above example: the apple which has been paid for can no longer be called "a gift").

However, Austinian and Husserlian phenomenology are still distinct and have a different emphasis: Austin concentrates on a systematic survey of related locutions (his analysis would be incomplete if he overlooked some typical way of speaking), while Husserl aims at a systematic survey of related phenomena (his analysis would be incomplete if he overlooked an essential aspect of these phenomena). If there are typical features of phenomena to which there corresponds no typical way of speaking in ordinary language—and the phenomenologists claim that there are such features—then the field of "linguistic phenomenology" is narrower than the field of phenomenology proper.

VI

To have a set of neatly defined and fixed concepts is the goal of phenomenological analysis and not its starting point. Husserl states:

It is misleading and radically perverse . . . to ask in the first beginning for terminologies of the kind used only in the end, when the final results of great scientific developments are definitively fixed."²⁶ Those who . . . ask for 'definitions' as in the 'exact sciences' . . . are still so truly beginners, that they have not yet understood the essential nature of phenomenology nor the method of work which it entails."²⁷

Ingarden mentions how the philosophers educated by the logicians of the Polish School could not understand this: "They claimed, that without such definitions it was impossible to have any conversation. Any attempt to pass this . . . demand

²² H. G. Gadamer, *Wahrheit und Methode. Grundzüge einer philosophischen Hermeneutik*, 2nd ed. (Tübingen, 1965), p. 444. Cf. the review by J. F. Staal in *Foundations of Language*, vol. 3 (1967), pp. 203–205.

²³ Cf. Husserl, *Ideen I*, pp. 160–163.

²⁴ Ingarden, *op. cit.*, p. 306. Cf. Sect. III. above.

²⁵ J. L. Austin, "A Plea for Excuses," *Proceedings of the Aristotelian Society*, vol. 57 (1956–1957), p. 8; A. J. Ayer, "Phenomenology and Linguistic Analysis," *ibid.*, supplementary vol. XXXIII (1959), pp. 111–124.

²⁶ Husserl, *Ideen I*, p. 206. Cf. also p. 9.

²⁷ Husserl, *Ideen I*, p. 207.

over . . . was countered with the famous reply "I do not understand . . ." ²⁸ Ingarden, who was very much frustrated by this refusal of communication, tried to show that the demand to have clear-cut definitions from the start was impossible and nonsensical. ²⁹

Actually I think that the logistic and the phenomenological procedures, which differ especially in this point, are both respectable. The logistic philosopher puts everything at once into concise and well-defined formulas, not because he thinks he has already got the final answer, but because he wants to check exactly what conclusions are implied. If unacceptable conclusions can be derived, then he will modify his "explication," i.e., the formulas in question, until the results are satisfactory. The phenomenologist on the other hand wants to improve his concepts in an intuitive way: he tries to formulate his thoughts in ordinary language or in a language close to ordinary language, because this language has a very rich vocabulary which helps us focus the attention on the phenomena in their concrete fullness. He checks intuitively whether his formulation "fits," i.e., whether it is "fulfilled" by his intuitive experience.

Both procedures, the logistic as well as the phenomenological, have their irreplaceable function. Where the intuitive grasp is weak or where we look for the systematic consolidation of our conceptual scheme and freedom from contradictions, logistic procedures are indicated. But the phenomenological explication is logically prior, because a minimum of intuitive inspiration is a prerequisite for any formal reconstruction. The more features a phenomenological kind of analysis uncovers, the more we can try to build into our logistic constructional system.

VII

Up to now we have concentrated on what the phenomenologist is doing all by himself. But the communication of one's insights to a colleague

presents another problem, whose difficulty is well known. As Ingarden points out:

It is impossible—once the ultimate results of our investigation have been obtained—to submit them to the reader exclusively by means of the new concepts which we have obtained on the basis of our phenomenological analysis, since in general the non-phenomenologist could in no way understand these concepts and would necessarily have to cast the phenomenological results into his own, different and inadequate conceptual apparatus. ³⁰

We seem to be faced with the impossible choice between using an adequate language which the reader does not understand and using a language which the reader does understand but which is inadequate for the expression of our findings.

However, there is a way out of this dilemma: ³¹ it is true that we cannot *directly* transmit the *outcome* of our analysis, but we can try to lead the reader step by step along the way which we have been going ourselves. By referring him constantly to the phenomena, we try to induce him to form the same concepts which *we* have in mind *now*. As knowledgeable guides we may even be able to help him avoid some traps into which we had fallen on our way, and we may be able to show him some shortcuts.

The starting point is again ordinary language. It provides a common ground for the phenomenologist and his reader, and it is best suited to bring the reader into live contact with the phenomena. Ingarden recommends to use at first strong, colorful words "taken not from books, but from life" even if they should be less precise than more technical terms. ³²

But it is the reader who has to make the effort of looking at the object and of trying to understand which features we want to point out to him. Ingarden insists very strongly that a reader who does not actively cooperate and refuses to make the effort of adapting his ordinary conceptual framework to the features of the phenomena pointed out to him, will always find the phenomenological writings ununderstandable. ^{33, 34}

University of Notre Dame

Received June 12, 1968

²⁸ Ingarden, *op. cit.*, p. 302 note.

²⁹ *Ibid.*, p. 301.

³⁰ *Ibid.*, p. 314. Cf. also Husserl, *Ideen I*, p. 206, where the author states that for the beginning the phenomenologist did well if he chose "imaginative expressions" (*bildliche Ausdrücke*), which were capable of directing his glance to clear instances of the phenomena, even if these expressions were less accurate than other more abstract terms.

³¹ Cf. Ingarden, *op. cit.*, pp. 315-317.

³² This paper was read on May 2, 1968, at the Western Division meeting of the American Philosophical Association in St. Louis.

²⁹ *Ibid.*, pp. 302-305.

³¹ *Ibid.*, p. 314f.

X. HUMAN RIGHTS AS ABSOLUTE CLAIMS AND REASONABLE EXPECTATIONS

J. E. BARNHART

HUMAN rights may be discussed under three broad headings: (1) access to freedom of self-fulfillment or happiness,¹ (2) fairness, and (3) compensation.

I

Every human need and desire, ideal and aspiration, interest and preference makes on the world its claim to satisfaction. In itself alone (i.e., abstractly), each claim stands absolute, above being either justified or unjustified. If there were no claims, the possibility of justification would not even occur.² In its isolated purity each claim is an absolute and an unqualified right. Of course, it is impossible to find absolute satisfaction for all claims. But *if* such a utopia were possible, it would not need to be justified. The significance of this point is that a genuine increase in over-all satisfaction does not need to be justified, for satisfaction of claims is that which makes the process of justifying possible and intelligible. Attempting to justify satisfaction *per se* is like attempting to give a straight answer to the question as to why there should be something rather than nothing.

Nevertheless, because in the real world claims do not occur in abstract and absolute isolation, the *possibility* of justification becomes the *necessity* of justification. Still it is utterly imperative to keep in mind that the question of the justification of each claim arises *only* in its relationship to the other claims. This is another way of saying that no claim is *a priori* justified over another. Justification is contextual or relational. Much of the history of the verbal conflict regarding human rights is the history of the practice of question begging and *a priorism*. Even H. L. A. Hart's "one natural right" of "freedom" is an example of how easily certain political claims (as over against economic, social,

etc., claims) are assumed to be the basis of human rights. (Cf. also Isaiah Berlin's "negative freedom.") Marxists, at the other extreme, in *a priori* fashion take the satisfaction of economic claims to be the essence of human rights. My point is that while satisfaction *qua* satisfaction is not a candidate for being justified or unjustified, each specific claim to satisfaction is a candidate for being evaluated with regard to its effects and influences on the other claims.

II

As Hobbes recognizes, the tragedy of finite human existence is that there is in the world no pre-established harmony of needs, ideals, and interests among men. If unqualified freedom to pursue one's own way were permitted as an actual right, life would be "nasty, brutish, and short." Given the social and natural conditions of man, individual freedom must be made relative. Furthermore, as social creatures, men cannot live apart from some cultural structure. Primates are not self-contained individualists who deliberately contract to work as a team only so long as they think the arrangement advisable. Men, like other primates, naturally live together in troops, groups, or communities. (With men, the forms which the community takes vary greatly.) To be human is to be a member of a community of some sort and complexity. Indeed, of all the animals, the human child is the most dependent upon extended family care. Lacking many of the instincts of the "lower" animals, man must have social structures and rules if he is to survive and enjoy even a limited self-fulfillment.

Born as an absolutist in his claim to freedom of self-fulfillment, the human child must be taught in some way (varying with the particular culture) that other persons make claims also. The autarchic

¹ No one ethical theory is defended in this article, although the arguments of the article best fit with some sort of ethic that emphasizes satisfaction of needs, desires, and interests.

² Cf. John Dewey, *The Quest for Certainty* (New York; Minton, Balch & Co., 1929), pp. 259f., 266, 268.

fiction of infancy becomes exposed to the fact that to live is to be involved with others. Without the "Thou," the "I" would not be. Rules and structures, as essential ingredients of any community, are a means of making it possible for members of the community to mitigate their absolute claims on each other, to aid each other in pressing their claims on the rest of their world, and to share in corporate values. Warriors, medicine men, farmers, politicians, scholars, engineers, merchants, etc., have all to some extent taken their separate roles as if the roles were instincts of the community.³ Hobbes, and later the Social Darwinists, failed to appreciate adequately the cooperativeness among men in the development of their interests. Anarchy as something other than a temporary stage in reorganization is a pure fiction, as Margaret MacDonald indicates.⁴

Now, by the mere fact of living with other members of the same species, men have found it necessary not only to restrain some of their interests and ideals but also to assume some kind of responsible role(s) through which to serve the community. Even non-human primates have ways of disciplining members who misbehave and who fail to take up their roles. Among baboons and macaques, for example, exists a "constant monitoring of gestures and the penalties for inattention by other troop members."⁵ With man, the ability to appreciate cue complexity and to be aroused by "sentiment structures" is extremely high.

The rules, regulations, and aspirations of the community must somehow be effectively communicated, otherwise the members cannot be expected to perform responsibly. This is why no one can meaningfully doubt that the community member has a *right* to expect the cues and signals to be communicated to him in some effective way. That is, the community member *rightly expects* to receive appropriate communications so that he may properly perform his roles and duties, and the community in turn *rightly expects* the individual to

carry out his roles so that the community may advance its claims upon its environment.

Perhaps it should be stressed that the concept of "fairness" means in part that the community member must be able to enjoy a practical measure of *consistency* and *clarity* regarding the community's expectations, cues, and signals. (In most human societies, clarity regarding some cues must come only after a long learning process.) If the rules, regulations, and expectations are not applied with consistency to the individual, then by definition unfairness emerges. The group expects to "count on" the individual to conduct himself circumspectly, and he in turn must be able to "count on" the community to act toward him in at least a consistent manner. (The law of "[Only] an eye for an eye" was a great leap in progress in human relationships. Arbitrariness was restrained considerably.)

Fairness in communication of expectations does not necessarily mean that all men will be treated equally and non-arbitrarily in every significant respect. But the principle of equality is at this stage implicit: The community must give each member—*regardless of his status and role*—a consistent and clear communication regarding what is expected of him. This consistency and clarity serves, of course, the advantage of the community, which, if it is to survive, must be able to expect certain services and responsible behavior from its members. This expectation may be called the community's right or claim upon the individual. T. H. Green and Margaret MacDonald are correct to expose the superficiality of the popular social contract theory of government. Nevertheless, it is possible to say that *implicit and functioning* within the reasonably successful community is something analogous to a social contract. A community may either deliberately or unconsciously revise or replace particular aspects of the "contract," but no community can hope to survive apart from some kind of implicit contract. (The fact that a kind of

³ Cf. Adam Smith's "invisible hand" and Hegel's "Cunning of Reason." Otters do not have to be taught to swim or catch fish. But turn a human child loose after he is weaned, and his chance of survival will be slim indeed. Lower animals live by certain "universal species habits" (instincts) that are genetically transmitted from parent to offspring. Man, however, seems to be comparatively limited in the number of automatic and physiological species habits he inherits. Hence, school teachers, parents, neighbors, etc. (or the equivalent) must teach the young how to respond in numerous situations. For example, respect for the claims of one's opponent to speak without his being shouted down or mobbed is not inherited genetically but is taught and demonstrated by junior-high school civics teachers, parents, etc., whose social roles are utterly essential to the life of the culture and the individuals. Of course, if the agents of the social roles are so inflexible as to ignore significant changes in the surroundings, their efforts may be compared to the harm resulting from "blind" instincts maladapted to unusual conditions. As a biological non-specialist, man's intelligence has been his major instrument of adaptation.

⁴ Cf. "The Language of Political Theory," *Proceedings of the Aristotelian Society*, vol. 41 (1940-1941), pp. 91-112.

⁵ R. L. Halloway, Jr., "Human Aggression: The Need for a Species-Specific Framework" in Morton Fried and Others (eds.), *War: The Anthropology of Armed Conflict and Aggression* (Garden City, N.J.; Natural History Press, 1968), p. 45.

social contract *functions* among some animals was apparently ignored by Hobbes and Locke.)

III

A social structure, while necessary for the individual's survival (at least), exacts a considerable toll from the individual. While the old romantic notion of the free and happy individual existing apart from society is no longer believable, it nevertheless remains true that the individual and community often frustrate as well as sustain each other. (The "invisible hand" easily becomes "sleight of hand.") As one anthropologist writes, "Human evolution has been the evolution of a paradox. The evolution of the brain and social structure, and symbol systems has also meant an increase in frustration and aggression."⁶

President John F. Kennedy once declared, "Ask not what your country can do for you, but what you can do for your country." The individual does owe a debt to his community. But it is equally true that the community owes a debt to the individual. *Compensation* rises out of freedom and fairness, and their frustration. There must be some form of mutual compensation between community and individual *if* each is to function successfully.⁷ This mutual compensation may even be called a mutual *natural* right in the sense that *compensation is necessary if men are to offset the frustration and aggression that seem to be natural to at least some kinds of social structures.*

Human rights are instrumental to human survival and to various styles of life. Whenever human right is conceived of as compensation, the question emerges regarding how much and what kind of compensation can be expected. The answer to this must depend upon what the natural and technological environment permits and how successfully individual freedom and fairness are used as guidelines for distribution. One curious trait of many oppressors is that they fail to consider

seriously that the oppressed do in fact make serious claims (in the sense of expressing needs, desires, ideas, and interests).⁸ The oppressor, then, is put on the defense to show why his claim should exceed or preclude that of the oppressed. The difficulty that the oppressor has, of course, is to avoid gross self-deception and the consequences that follow. For example, the Nazis apparently realized that to declare boldly that Jews and non-Aryans ought to be oppressed merely because they are Jews and non-Aryans, would be to set forth a "principle" of wholesale arbitrariness so drastic in its implications as to destroy all grounds for loyalty and unity among even the Nazi Aryans. Hence, the Nazis had to indulge in massive self-deception. They convinced themselves that Jews are vile and wicked creatures, that Jesus of Nazareth was not a Jew, and that Gentile non-Aryans are stupid and weak. Indeed, Hitler's fantastic belief in the superiority of Aryan Germans allowed him to believe that they could simultaneously destroy Russia and England with little difficulty. Toward the end of his life, Hitler came to despise and hate real Germans who failed to measure up to his crazy image of them as an invincible super race.

Similarly the white racist in the United States has resorted to lying to himself about his own innate superiority over Negroes. In attempting to live up to this image, he has been forced to turn his back upon fact after fact. He has had to deny Negroes opportunities so that he could "prove" that Negroes could "never amount to much." He has made prophecies about the failures of Negroes, and then has had to take measures to guarantee that the Negroes would not be given opportunities to succeed. The white racist has now come to the point of beginning vaguely to feel that *he himself* must be the most stupid of all men as he watches many Negroes succeed despite white bigotry and overwhelming disadvantages.⁹ What grinds in the craw of the white racist in the United States is the

⁶ Holloway, "Human Aggression . . .," *op. cit.*, p. 47. Freud notes that often love and hate develop together.

⁷ The social contract as a conceptual model calls attention to the fact that men *need, expect, and demand* certain things from one another. It also suggests that a certain cultivated *agreeableness* is necessary if the individuals of a community are to develop economic, religious, literary, etc., interests. In many respects the pre-Darwin social contract school apprehended vaguely at least the fact which modern anthropologists more fully understand, namely, that "the emergence of human society required some suppression, rather than a direct expression, of man's primate nature" (M. D. Sahllins, "The Origin of Society," *Scientific American*, vol. 203 [1960], p. 77). "Gratitude," says John Dewey, "is the root of all virtue" (*Human Nature and Conduct* [N.Y., Modern Library, 1930], p. 21). The social contract model is a kind of self-conscious recognition of citizens' mutual indebtedness to one another.

⁸ White segregationists have pretended that Negroes in particular make very few claims. This is why segregationists hated the articulate Martin Luther King, Jr., so deeply: He exposed their fabrication.

⁹ Many white racists who identify manhood with sexual prowess have long believed that the Negro's sexual organ is larger than the white man's.

horrible thought that he himself is the inferior member of the species. In desperation, like a rationalizing true believer, he sees himself as having been victimized by a conspiracy of Jews and Communists. Paranoia and self-hatred run rife among modern white racists, who fight blindly against the fear that they themselves may be sinking to the bottom of the pecking order.

There are many undesirable consequences which the oppressor must learn to face. In the South of the United States many small communities have "allowed" the KKK to carry out threats and foul deeds on their behalf. The whites in some of those communities have now begun to realize that they themselves are being intimidated by the KKK.

The point to be made is not that crime never pays or that goodness always triumphs. Rather, it is that oppressors cannot enjoy their special privileges without paying dearly for them. To use the phrase "human right," then, is often a way of reminding the oppressors that they after all are oppressors (oppressors have a habit of thinking themselves to be benefactors!) and that their way of life carries special consequences which they may not have faced realistically. It is interesting that the Civil Rights movement in the United States may be on the verge of moving toward Black Power, power even in the sense of physical force and violence. This power which the black man is learning to exert was learned to a great extent from his white oppressors. The price which white Americans will pay to maintain their racism may be more than was counted on.

IV

In the United States it is often said that every citizen has the right to the "pursuit of happiness." That is, he may seek to gain happiness so long as he does not violate the laws. Among the right-wing ideologists in the United States is the belief that while all citizens have the right to *pursue* happiness, none has a right to happiness itself. However, this belief fails to call attention to the fact that the pursuit of happiness (to use a popular phrase which needs to be analyzed) does not take place in an ideal setting. If there is good reason to believe that a measure of happiness could not in fact be attained in the *actual* setting, then to talk of the right to *pursue* happiness is misleading. If a

man in prison is told that he has a right to have a courtship and subsequently to enjoy married life, he must assume that he will either be released from prison so that he may pursue marriage or be allowed to associate in prison with at least one woman. Similarly, to tell a person that he has the right to pursue happiness implies that the recalcitrant circumstances are not so great as to make the pursuit practically impossible. "Human rights," then, are claims and expectations which the individual rightly and reasonably makes within the social, ecological, and technological context.

An individual cannot rightly make a claim on a community that is without the means to meet the claim. It is wrong to make such a claim for the simple reason that it cannot be met. On the other hand, if a community comes to believe that, say, dental care is possible for all citizens and essential for the happiness of average (middle class) citizens, then each citizen has a right to dental care. How this right or claim is to be carried out may vary greatly; but if those below the middle class are denied available means to dental care, then their right has been trampled upon. As R. B. Perry notes, few things are so cruel as giving to needless neglect the name of "liberty."¹⁰ If a middle-class citizen protests that he himself worked and thus earned money as a means to satisfying his right to dental care, he is implying that others ought also to enjoy the opportunity to earn money to pay for dental care. (And so on regarding food, housing, clothing, education, etc.)

What is implied in the idea of "human rights" is the insistence that because all members of the community are expected to conform to the rules, regulations, and aspirations of the community, there must exist a minimal limit of means, status, opportunities, and liberty (i.e., whatever makes for moderate freedom and happiness) below which the deprived will be given every opportunity to prevent themselves from falling. Implied also is the elimination of an avoidable lower class existence—except in cases where individuals genuinely and freely choose to be lower class. And in some of these cases, the community may choose to prevent the self-made lower-class individuals from falling below a certain level of ignorance of certain forms and from contracting diseases, since such diseases and forms of ignorance create excessive hardship for the other members of the

¹⁰ Utilitarian Jeremy Bentham suggests that laws ought to be passed against those who, "without prejudicing" themselves, could assist someone in dire need but refuse to do so (cf. Bentham, *Introduction to the Principles of Morals and Legislation*, I, sect. 19 and note; also sect. 4).

community, including the offspring of the lower class.

A community that both denounces the lower class because of its deficiencies and denies it the opportunities to improve itself is plainly renouncing fairness and compensation, and thus is reducing excessively the *freedom* of the individuals of the lower class. (It is an upper-and-middle-class bias to regard freedom as nothing but political liberty.)

The discussion, production, and distribution of human rights is as open-ended as is human nature itself. Men have a great variety of needs, wants, desires, ideals, dreams, and preferences. Therefore, there is no sacred and fixed number of rights. So long as men make claims and are such that they must live with and be involved with one another, they must give some attention to fairness, compensation, and the determination of the limits and extents of freedom of self-fulfillment.

V

A final comment needs to be made regarding the problem of individuals and groups of persons who become useless to the economic market of the community. Do they cease to have rights? This problem is not simply that of, say, Eskimos having to dispose of elderly parents in order to permit the entire community to continue in its quest for survival. Rather, it is the special problem of a highly technological society in which many human beings at present—and the number will probably increase sharply—are utterly useless to the economy. In the United States today millions of blacks and whites are an economic surplus. Their ancestors may have helped make the economy what it is today, but the offspring no longer have market-roles. The community asks nothing from them except that they obey the laws

and keep out of trouble. Some Americans appear to think that if these roleless people were simply put aside in some efficient and quiet manner (for, after all, they do absorb consumer goods), life would be better for all concerned—the roleless are apparently regarded as of no concern.

One way of correcting this way of looking at the people without roles in the economy is to point out that the economic role is not the most important role, not today in a highly technological nation. Indeed, because of this economic role, many other roles have been denied development. Secondly, the concept of *work* must be made to become considerably more flexible. The early puritans would have denounced today's professional ball players as idle pleasure-seekers, not diligent workers. If the concept of work expands markedly, then completely new kinds of "jobs" may develop despite cybernation; for, after all, there will always be human problems to solve and human interests to pursue. But the desperate need is to guarantee that the new jobs will be regarded as contributing some value to the community. Furthermore, every citizen must be guaranteed the means for a decent living and the opportunities to live happily, lawfully, and in greater freedom.

The society that can bring itself to eliminate Negroes today will eliminate the white ex-factory worker tomorrow, the old person next week, the pesty child next month, and the eccentric neighbor next year. To use the phrase "human rights" is to call attention to the fact that cybernation was made (or ought to be made) for man, and not man for cybernation. To restate Bentham somewhat, the question is not "Of what economic use is a person?" but "Can he dream dreams, have interests, experience pleasure, respond to other humans, gain satisfaction, make plans, and enjoy some peaceful way of life?"

XI. CORRIGENDA

Volume 6 (1969)

GREGORY SHERIDAN: *The Electroencephalogram Argument Against Incurability* (pp. 62-70).

Page 63, col. 2, line 9 sentence should read: In both cases the apprehension of the occurrence will have to . . .

AMERICAN PHILOSOPHICAL QUARTERLY

MONOGRAPH SERIES

Edited by NICHOLAS RESCHER

The *American Philosophical Quarterly* also publishes a supplementary Monograph Series. Apart from the publication of original scholarly monographs, this is to include occasional collections of original articles on a common theme. These monographs can be purchased separately but are available to individual subscribers at *half price* (though not to institutional subscribers).

No. 1. STUDIES IN MORAL PHILOSOPHY. *Contents:* Kai Nielsen, "On Moral Truth"; Jesse Kalin, "On Ethical Egoism"; G. P. Henderson, "Moral Nihilism"; Michael Stocker, "Supererogation and Duties"; Lawrence Haworth, "Utility and Rights"; David Braybrooke, "Let Needs Diminish That Preferences May Prosper"; and Jerome B. Schneewind, "Whewell's Ethics." 1968, \$6.00.

No. 2. STUDIES IN LOGICAL THEORY. *Contents:* Montgomery Furth, "Two Types of Denotation"; Jaakko Hintikka, "Language-Games for Quantifiers"; James W. Cornman, "Types, Categories, and Nonsense"; Robert C. Stalnaker, "A Theory of Conditionals"; Alan Hausman and Charles Echelbarger, "Goodman's Nominalism"; Ted Honderich, "Truth: Austin, Strawson, Warnock"; and Colwyn Williamson, "Propositions and Abstract Propositions." 1968, \$6.00.

No. 3. STUDIES IN THE PHILOSOPHY OF SCIENCE. *Contents:* Peter Achinstein, "Explanation"; Keith Lehrer, "Theoretical Terms and Inductive Inference"; Lawrence Sklar, "The Conventionality of Geometry"; Mario Bunge, "What Are Physical Theories?"; B. R. Grunstra, "The Plausibility of the Entrenchment Concept"; Simon Blackburn, "Goodman's Paradox"; Stephen Spielman, "Assuming, Ascertaining, and Inductive Probability"; Joseph Agassi, "Popper on Learning from Experience"; D. H. Mellor, "Physics and Furniture"; and Michael Slote, "Religion, Science, and the Extraordinary." 1969, \$6.00.